The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

SMR/1849-12

Conference and School on Predictability of Natural Disasters for our Planet in Danger. A System View; Theory, Models, Data Analysis

25 June - 6 July, 2007

Predictability of Weather
Large Atmospheric Models
Parts I, II, IIIa, IIIb

Davis Straus

George Mason University
Calverton, USA

Strada Costiera 11, 34014 Trieste, Italy - Tel. +39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it, www.ictp.it

Predictability of Weather
Large Atmospheric Models

Part I: Physical Interpretation of Error Growth
(or loss of predictability) in the atmosphere in terms of:
Fundamental Linear Instabilities of Atmospheric Flow

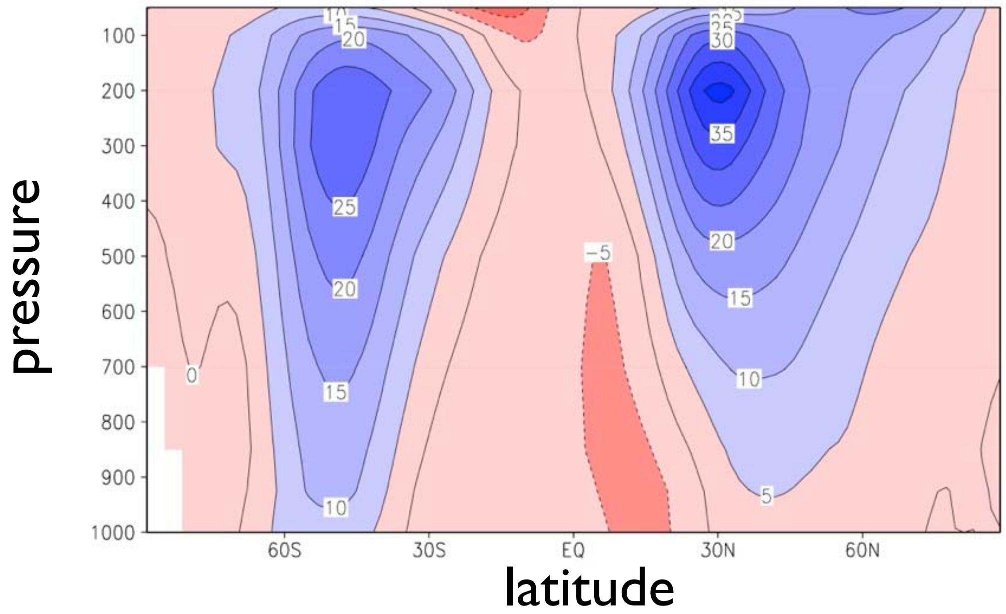
Part II: Error Growth in Numerical Weather Prediction Models

Part III: Scale Dependence of Error Growth in Large Models

Predictability of Weather
Large Atmospheric Models
Part I

Physical Interpretation of Error Growth
(or loss of predictability) in the atmosphere in terms of:
Fundamental Linear Instabilities of Atmospheric Flow

$p = p_s e^{-(z/H)}$

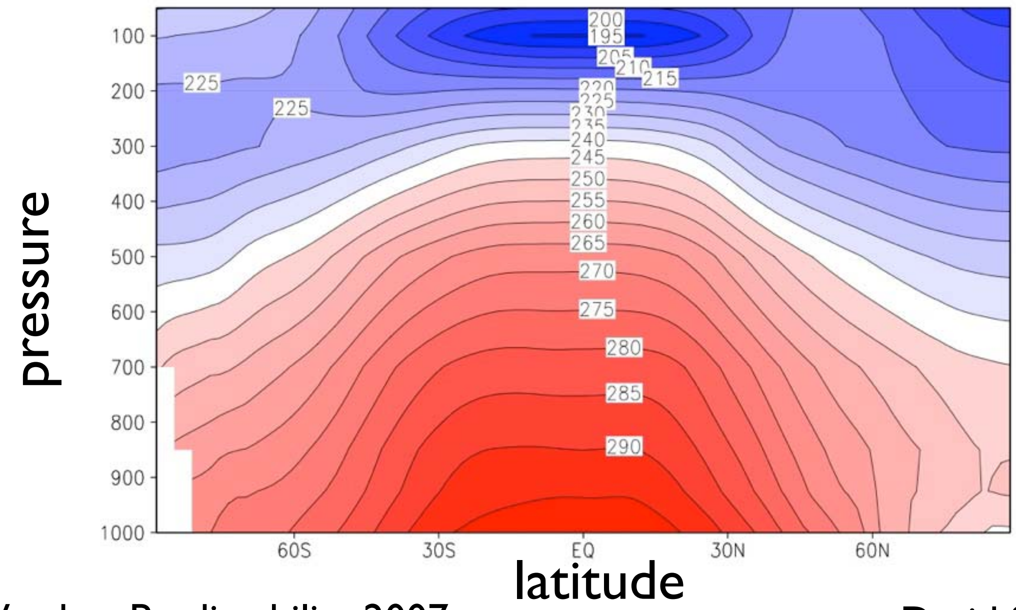


East-to-West "zonal" wind u (averaged over longitude)

Note the strong (positive) vertical shear in mid-latitudes

$$\frac{\partial u}{\partial z}$$

The seasonal (DJF) mean "basic state" of the troposphere



Temperature T (averaged over longitude)

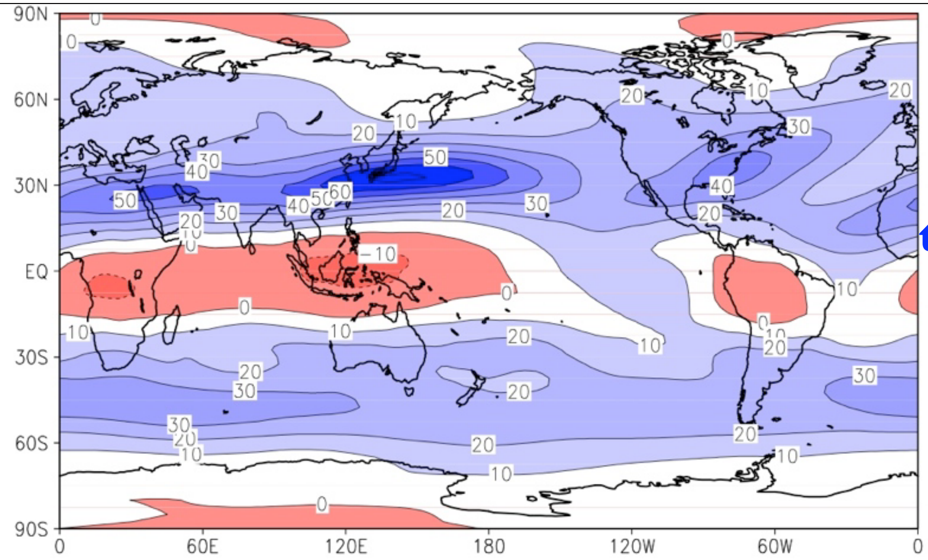
Note the strong negative poleward gradient of T in mid-latitudes

$$\frac{\partial T}{\partial y}$$

Localization of vertical shear in "jets" over the Pacific and Atlantic Oceans in winter

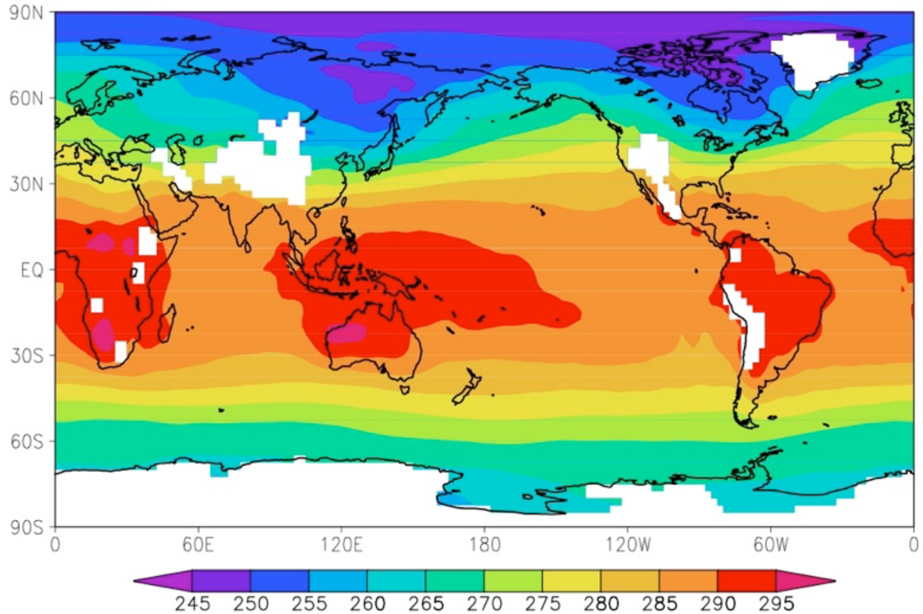
Presence of

$$\frac{\partial u}{\partial y}$$



Zonal wind u at the top of the troposphere (200 hPa)

The seasonal (DJF) mean "basic state" of the troposphere



Temperature T just above the boundary layer (850 hPa)

The presence of strong gradients in the mean flow

Baroclinic $$\frac{\partial u}{\partial z} \propto -\frac{\partial T}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y}$$ Barotropic

leads to the rapid growth of small wavy perturbations.
For small amplitudes of waves, this is a *linear* problem.
The study of these perturbations, and how their growth is related to the features of the gradients of the mean flow, is an important subject in geophysical fluid dynamics.

These wavy perturbations ultimately give rise to both weather fronts and cyclones in mid-latitudes, by a complex set of linear and non-linear interactions.

<u>Connection to Predictability (sensitivity to initial conditions):</u>

Consider the rapidly growing solution y(x,y,p,t) whose structure is fixed but whose amplitude grows with time:

$$y = y_0 e^{a(t-t_0)}$$

Since y_0 is the initial condition (value of y at $t=t_0$), any change in in the initial condition grows exponentially

Idealized Structure of baroclinic growing waves and their role in the general circulation

(1) They consist of waves that propagate eastward with periods of about 2 - 8 days

(2) They can be broken into components, each of which has a give zonal wavenumber m:

$$\psi = Ae^{i(m\lambda - \omega t + \Phi)} e^{\alpha t}$$

where the amplitude A and phase Φ are functions of (λ, φ, p), and the growth rate α and frequency ω are functions of the zonal wavenumber m and the basic state on which the wave propagates. (Note: Here λ is longitude and φ latitude). The streamfunction ψ is related to the horizontal flow by:

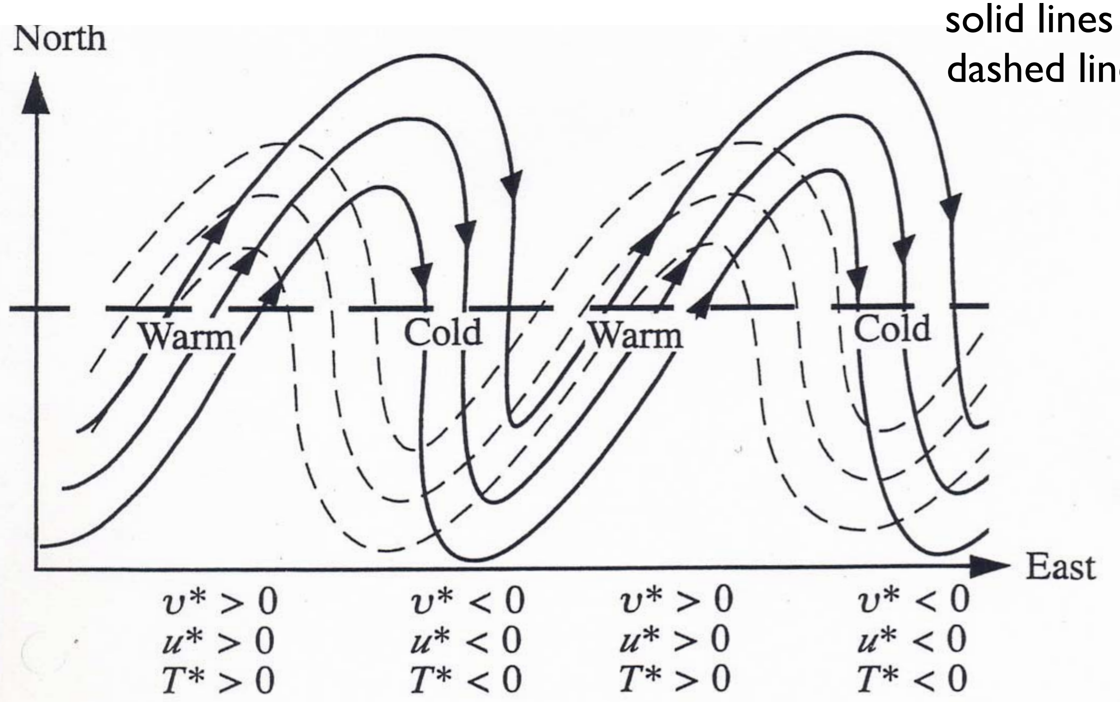
$$u = -\frac{1}{a}\frac{\partial \psi}{\partial \lambda}$$

$$v = +\frac{1}{a}\frac{\partial \psi}{\partial \phi}$$

where a is the earth's radius.

(3) Their vertical structure, and the relationship between temperature T and height (Z) or streamfunction (ψ) is such that they transport heat and momentum towards the pole.

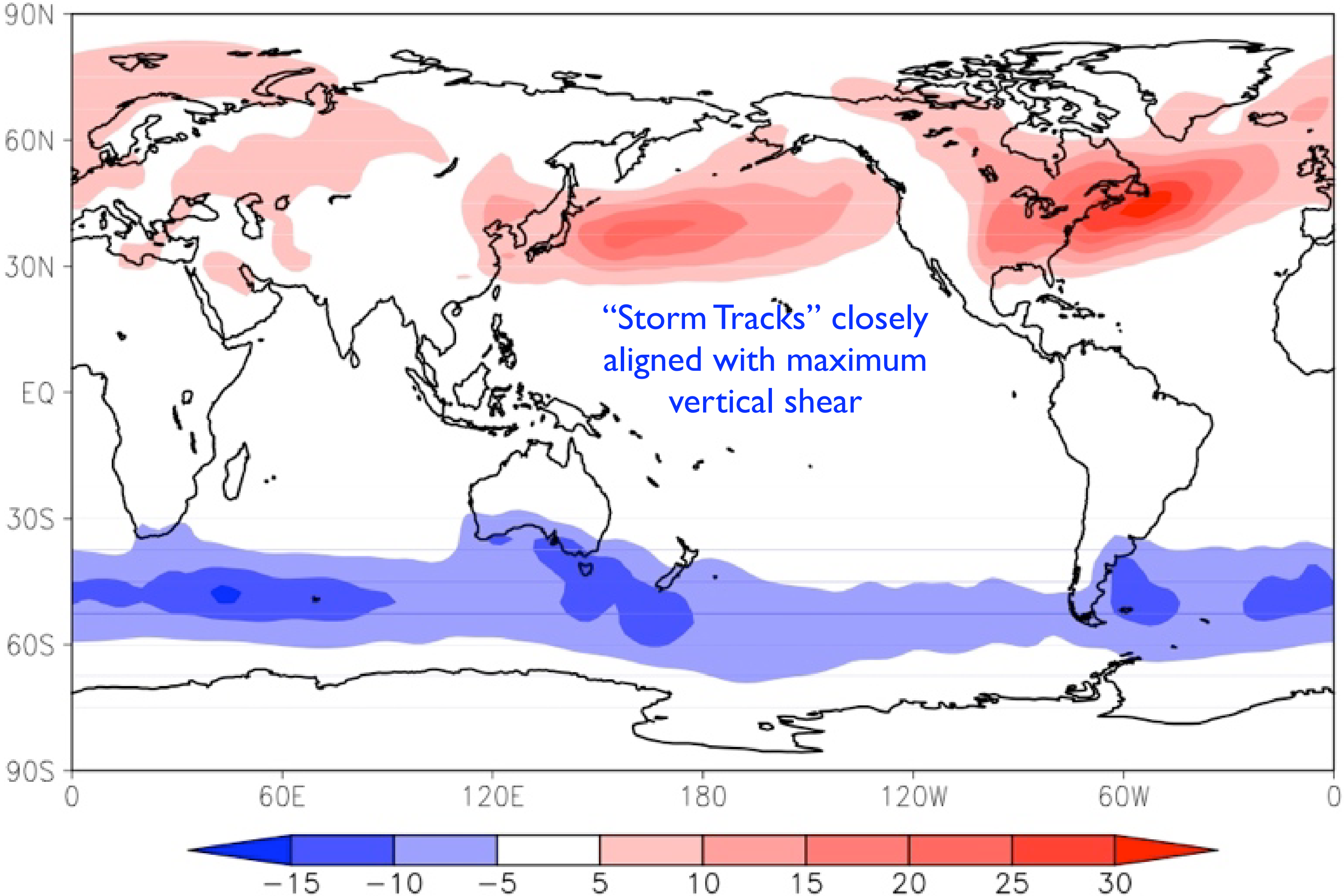
solid lines are streamfunction
dashed lines are temperature



Schematic of the streamlines (solid) and isotherms (dashed) associated with a large-scale .sturbance in midlatitudes of the Northern Hemisphere. Arrows along the streamline con- ıe direction of wind velocity. The streamlines correspond approximately to lines of constant the winds are nearly geostrophic. The signs of the deviations of the wind components from ːrage values are shown to illustrate that the NE–SW tilt of the streamlines indicates a north- ɔmentum transport, and the westward phase shift of the temperature wave relative to the gives a northward heat transport.

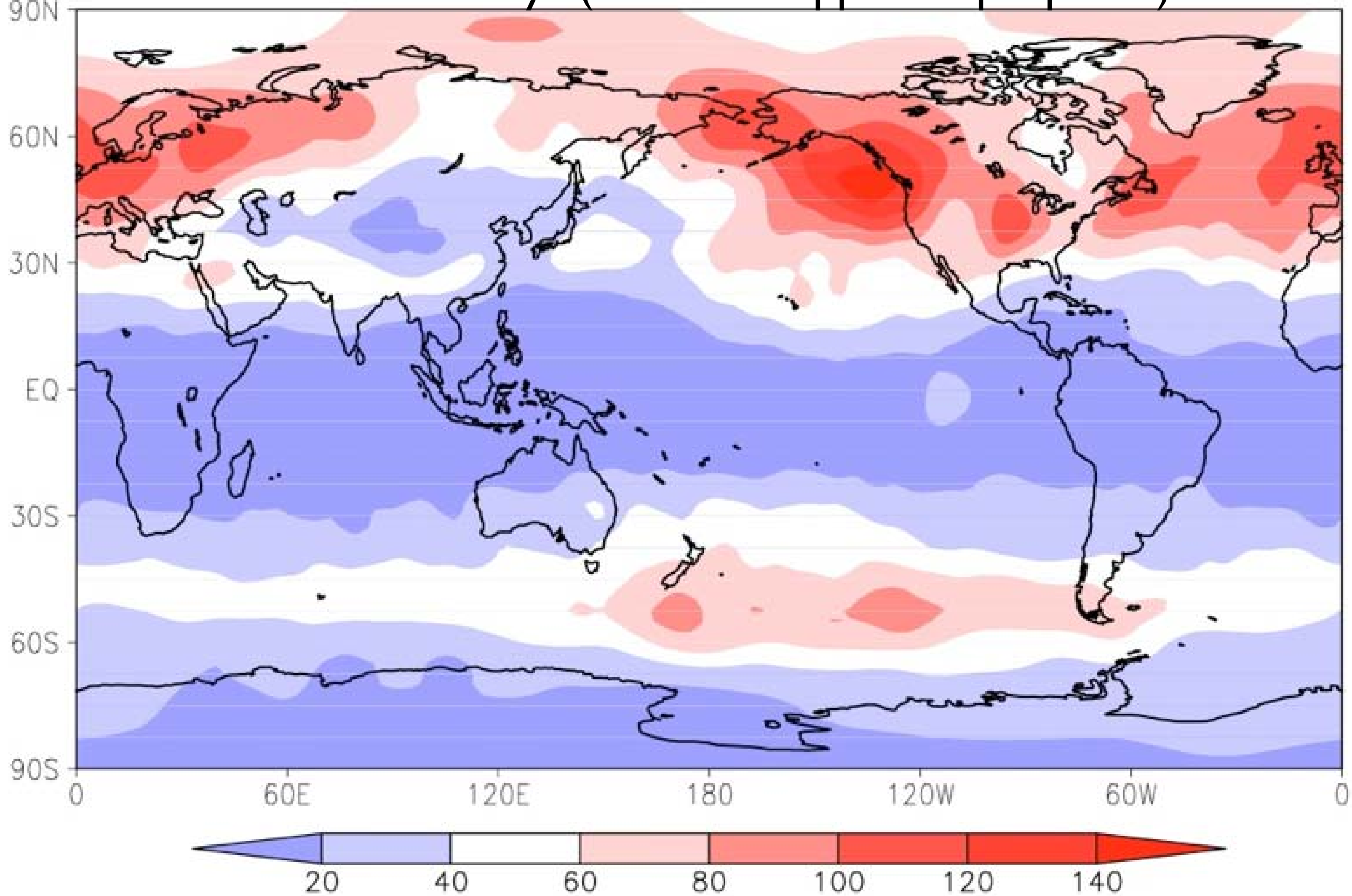
from: "Global Physical Climatology" by Dennis Hartmann
(Academic Press, 1994)

Covariance of northward velocity (v) and Temperature T due to fluctuations with time scales of 2-8 days $\overline{v'T'}$



Physically: Poleward heat transport by waves that form due to baroclinic instability. The waves try to smooth out the temperature gradient

Variance of northward velocity (v) due to fluctuations with time scales of 8-90 days (300 hPa - upper troposphere)



Physically: Maxima related to “blocking highs” , which are long-lived high pressure systems that cause persistent weather regimes downstream

Types of Large Numerical Weather Models:

(1) Atmospheric General Circulation Models (Numerical Weather Prediction Models or Climate Models)

(2) Turbulence models (idealized) - two-dimensional, quasi-geostrophic.
- Focus here is on wave-wave interactions, also called non-linear interactions
- Models typically cover many orders of magnitudes of scales by solving equations with “closure” schemes - a completely different philosophy of solving the equations
- Possible only for relatively simple governing equations

Atmospheric General Circulation Models

(1) Fundamental equations are the “primitive” equations

(2) parameterizations for processes not explicitly resolved (see below)

(3) Horizontal domain is global, with spherical geometry

(4) Vertical Domain encompasses the troposphere and stratosphere

(5) Difference between numerical weather prediction models and climate models are primarily resolution

Primitive Equations

(1) Filtered version of the fundamental equations for fluid dynamics "Navier-Stokes" equations.

(2) Assumption made that vertical domain is much smaller than horizontal domain, so that the vertical velocity is much smaller than the horizontal velocity.

(3) Assumption (2) is consistent with the hydrostatic equation, which relates the mass to the vertical derivative of pressure. This filters out sound waves from the set of equations.

(4) Often solved with the use of pressure (or a related quantity) as the vertical coordinate. In these "pressure" coordinates, the fundamental dynamical equations consist of:

Momentum equations for horizontal flow - Newton's Second Law in a rotating frame of reference. ($F = ma = m\, dv/dt$)

Thermodynamic equation ($T\, dS/dt = Q$) where S=entropy, Q=heating

Conservation of mass

Conservation of water (vapor + liquid + ice)

Primitive Equations (continued)

(5) In these fundamental equations the basic non-linearity arises because of the distinction between Lagrangian rates of change (following a parcel of air), and Eulerian rates of change, (expressed) in a fixed (x,y,p,t) coordinate system:

(Lagrangian) $$\frac{dT}{dt} = \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + \omega\frac{\partial T}{\partial p}$$ (Eulerian)

(6) The fundamental dynamical equations of motion are supplemented by very important physical processes, a few of which are listed here.

Radiation - both incoming and reflected solar radiation, and thermal radiation from the ground, from gases in the troposphere, and from clouds.

Latent heat release from condensation of water vapor due to resolved motions.

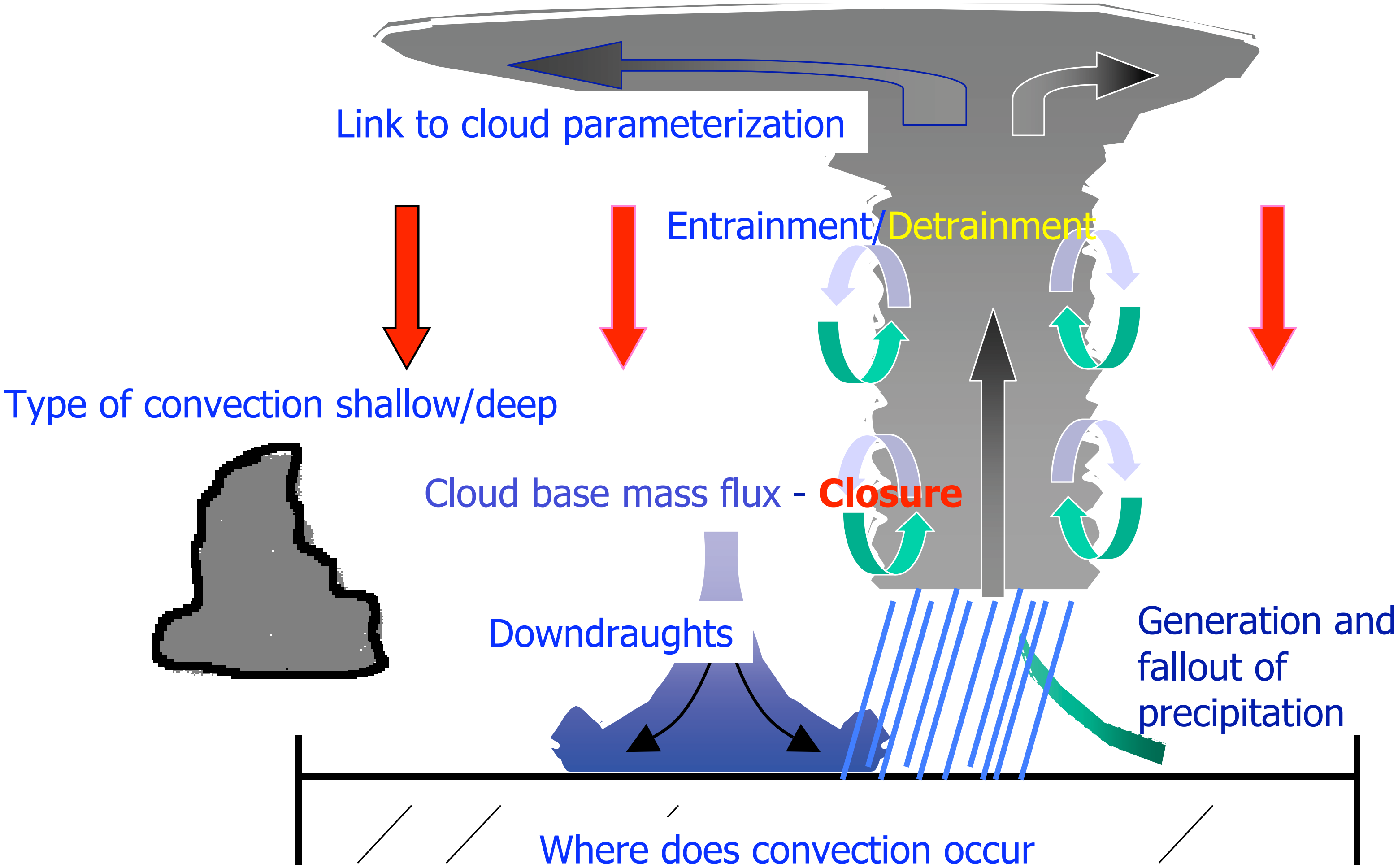
Latent heat release and motion due to motions not resolved - convection (includes both deep convection and shallow convection).

Planetary Boundary Layer diffusion and turbulence.

These processes can be very non-linear, if fact not even analytic!

A bulk mass flux scheme:

What needs to be considered



(Andreas Chlond)

Atmospheric Forecast and Climate Models

Variables which predicted by explicit evolution equations ("prognostic"):

At every model vertical level:
(1) Horizontal flow (zonal and meridional winds)
(2) Temperature
(3) Water Vapor
(4) Liquid and Solid water (clouds)

Surface Fields:
(1) surface pressure
(2) wind stress on the ocean

Total number of variables (estimated from NCEP Global Forecast System):
http://wwwt.emc.ncep.noaa.gov/gmb/moorthi/gam.html

Weather forecasting configuration ~ 47,100,000
Climate forecasting configuration ~ 1,300,000

Very rough estimate of smallest scale resolved:

Weather forecasting configuration ~ 100 km
Climate forecasting configuration ~ 650 km

Predictability of Weather
Large Atmospheric Models
Part II

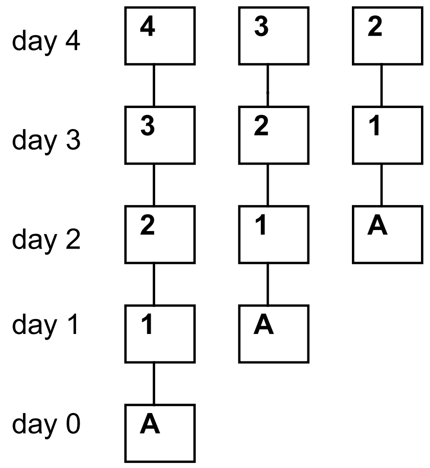
Error Growth in Numerical Weather Prediction Models

Weather Predictability:
Growth of Errors in Operational Forecast Models

Lorenz, E. N., 1982: Atmospheric Predictability Experiments with a Large Numerical Model. *Tellus,* **34,** 505-513.

Dalcher, A., and E. Kalnay, 1987: Error Growth and Predictability in Operational ECMWF Forecasts. *Tellus,* **39A,** 474-491.

Simmons, A., and A. Hollingsworth, 2002: Some Aspects of the Improvement in Skill of Numerical Weather Prediction, *Quart. J. Roy. Meteor. Soc.,* **128,** 647-678.



Schematic of Analysis Forecast System. The bold letter A within a box refers to an analysis for the given day, that is the estimate of the real state of the atmosphere on that day. The bold numbers in the boxes refer to the range of the forecast. The days labeled on the left refer to the verifying time

Analysis and Forecast

Analysis is the estimation of the current state of the atmosphere, expressed as a state of a numerical model (denoted by "A" in diagram).* The analysis is expressed in terms of all model prognostic variables, on the model's horizontal and vertical grid.

Forecast is a projection into the future made using a numerical model from an initial state given by the Analysis.

*** In practice, the analysis starts from a previous very short range (6 hour) forecast - the model variables are changed to be consistent with the current observations for those areas/levels/ variables which are observed**

The difference between a forecast which has been run for N days and the analysis at the end of the N days (the so-called "verifying analysis") is called the forecast error.

Forecast error has several components:
(1) analysis error: The initial conditions obtained from the analysis may have errors that are not small.
(2) model error: The model itself has physical errors.
(3) predictability error: Any (inevitable) small errors in the analysis will amplify with time.

Definition of predictability error:
The difference between two model forecasts started from initial conditions very close to each other. (The predictability error measures the forecast error that would be seen if the models were perfect and the analysis very good.)

Identical twin model configuration:
Run forecasts with the same model, but with initial states close to each other.

It is generally believed that with current NWP forecast models (e.g. ECMWF), the forecast error *for the first few days* is dominated by the analysis error.

Operational ECMWF forecasts and analyses for winter of 1980 / 1981

100 - day period starting from 1 December 1980

For each day we have the analysis, and forecasts starting from that analysis for 1 day, 2 days, ... 10 days.

Consider measure of error between two forecasts at any time to be the square root of the global mean of the squared difference between 500 hPa height (Z). This error is called the rms error.

Consider two forecasts one started j days earlier, one started k days earlier, both verifying for (valid for) the current time.

Example:
Forecast A is a 2-day forecast started on 23 December (j=2)
Forecast B is a 3-day forecast started on 22 December (j=3)

The "error" in height between these two forecasts valid for Dec. 25 is:

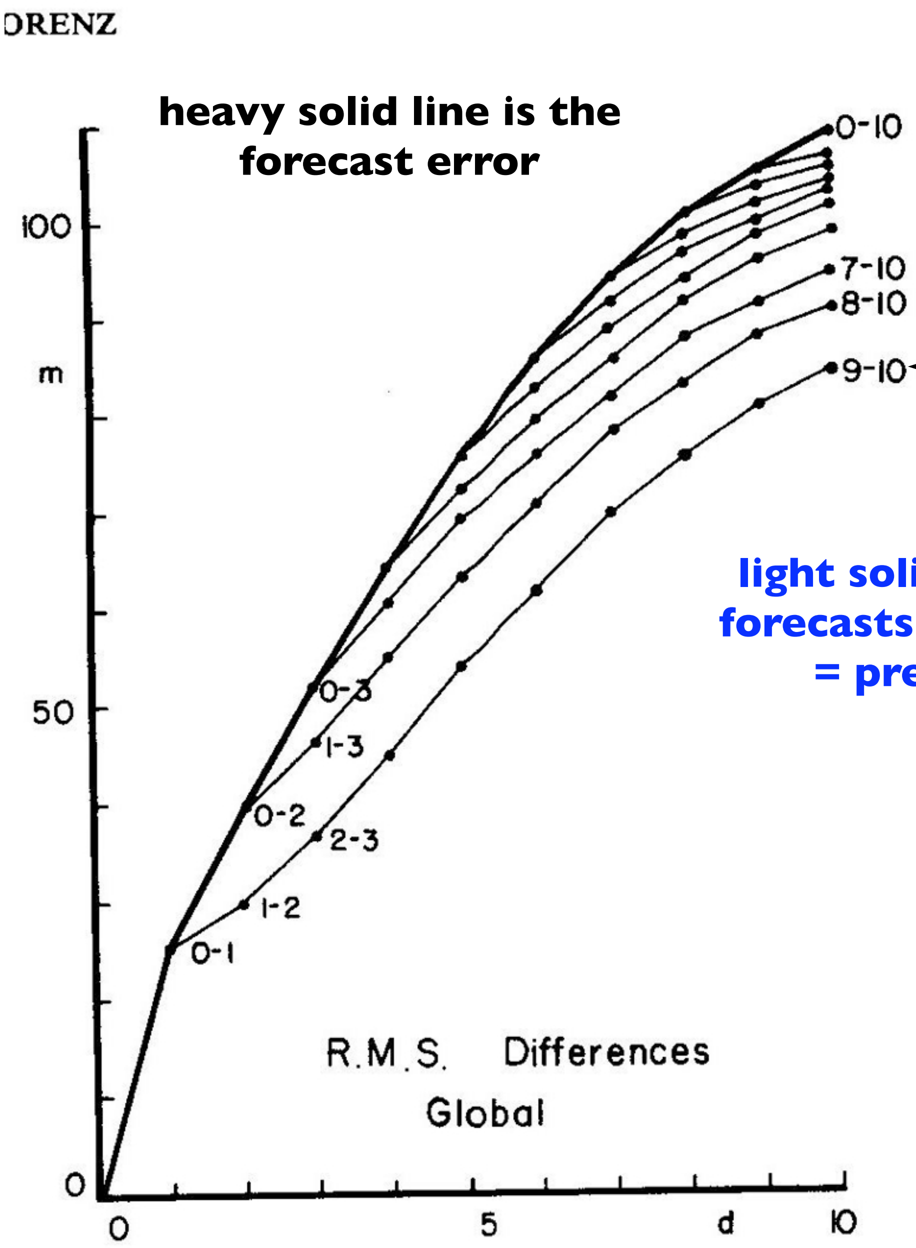
$$E_{j,k}^2 = \frac{1}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\phi \int_0^{2\pi} d\lambda \cos(\phi) \ (Z_A - Z_B)^2$$

where ϕ is latitude and λ longitude.

$E_{(j,k)}$

j=0 is the analysis

heavy solid line is the forecast error



Focus on this curve

light solid lines are errors of forecasts started d days apart = predictability error

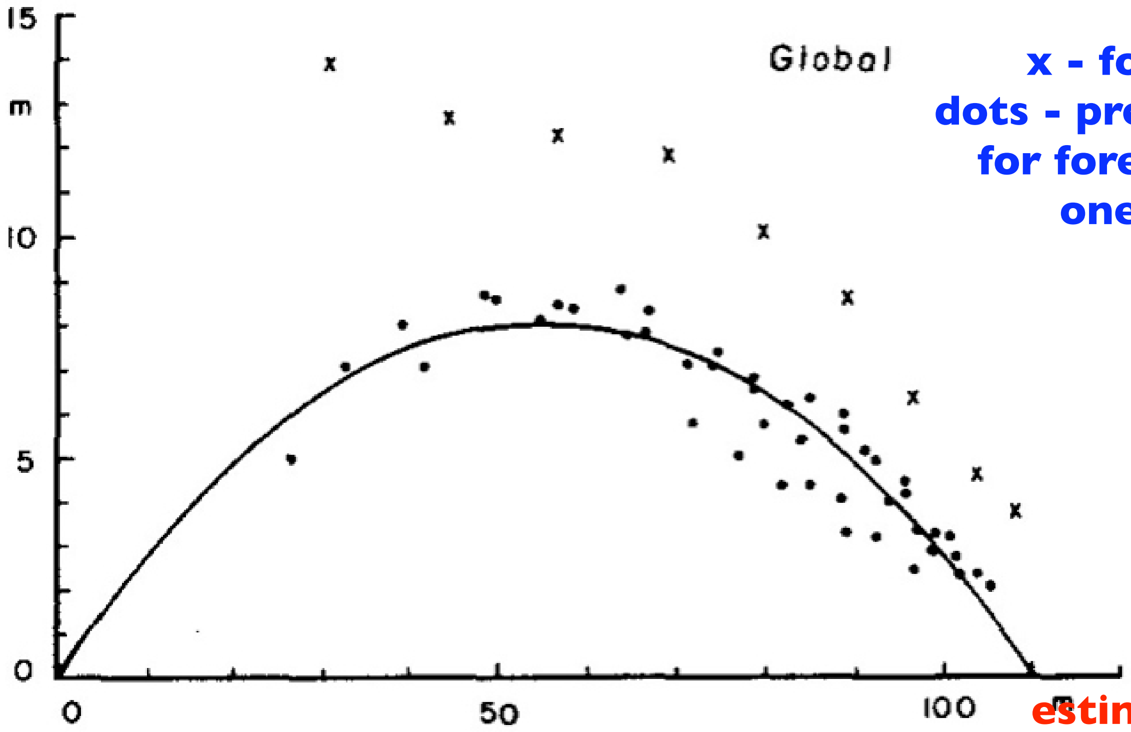


Fig. 2. Increases in global root-mean-square 500-mb height differences, $E_{j+1,k+1} - E_{jk}$, plotted against average height differences $(E_{j+1,k+1} + E_{jk})/2$, in meters, for each one-day segment of each thin curve in Fig. 1 (large dots), and increases $E_{0,k+1} - E_{0k}$ plotted against average differences $(E_{o,k+1} + E_{0k})/2$, for each one-day segment of heavy curve in Fig. 1 (crosses). Parabola of "best fit" to large dots is shown; see text.

previous figure suggests a parabolic form for error growth:

$$\frac{dE}{dt} = aE - bE^2$$

E = rms of 500 hPa z

$$Lim_{t \to \infty} E = E_\infty$$

At large time, errors should saturate

$$aE_\infty = bE_\infty^2 \qquad E_\infty = a/b$$

$$\frac{d\varepsilon}{dt} = a\varepsilon - bE_\infty \varepsilon^2 = a\left(\varepsilon - \varepsilon^2\right) = a\varepsilon\left(1-\varepsilon\right) \qquad \varepsilon \equiv E/E_\infty$$

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = \left(\frac{1}{1-\varepsilon}\right)\frac{d\varepsilon}{dt} + \frac{\varepsilon}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt} = \frac{1}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt}$$

or

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = a\frac{\varepsilon}{1-\varepsilon}$$

defining: **we have**

$$f \equiv \frac{\varepsilon}{1-\varepsilon} \qquad \frac{df}{dt} = af \qquad f = e^{a(t-t_0)}$$

(here t_0 has a specific meaning)

$$\varepsilon = \frac{f}{1+f} = \frac{e^{a(t-t_0)}}{1+e^{a(t-t_0)}} = \frac{e^{\frac{1}{2}a(t-t_0)}}{e^{-\frac{1}{2}a(t-t_0)} + e^{\frac{1}{2}a(t-t_0)}}$$

t_0 is the time at which E = 1/2

or

$$\varepsilon = \frac{1}{2}\left[1 + \tanh\left(\frac{1}{2}a(t-t_0)\right)\right]$$

Some identities:

$$tanh(x) \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$1 + tanh(x) = \frac{2e^x}{e^x + e^{-x}} \qquad \frac{1}{2}(1 + tanh(x)) = \frac{e^x}{e^x + e^{-x}}$$

for $\varepsilon \ll 1$

we have $$\frac{d\varepsilon}{dt} = a\varepsilon$$

or $$\varepsilon = ce^{a(t-t_0)}$$

the doubling time for small errors τ **is defined by:**

$$\frac{\varepsilon(t_2)}{\varepsilon(t_1)} = e^{a(t_2-t_1)} = 2$$

or

$$(t_2 - t_1) \equiv \tau = \ln(2)/a$$

doubling time for small errors is 2.42 days

Modified model: Correct all forecast height fields (after the fact) so that at each grid point:
(1) mean over all forecasts will equal the observed mean
(2) variance over all forecasts will equal the observed variance

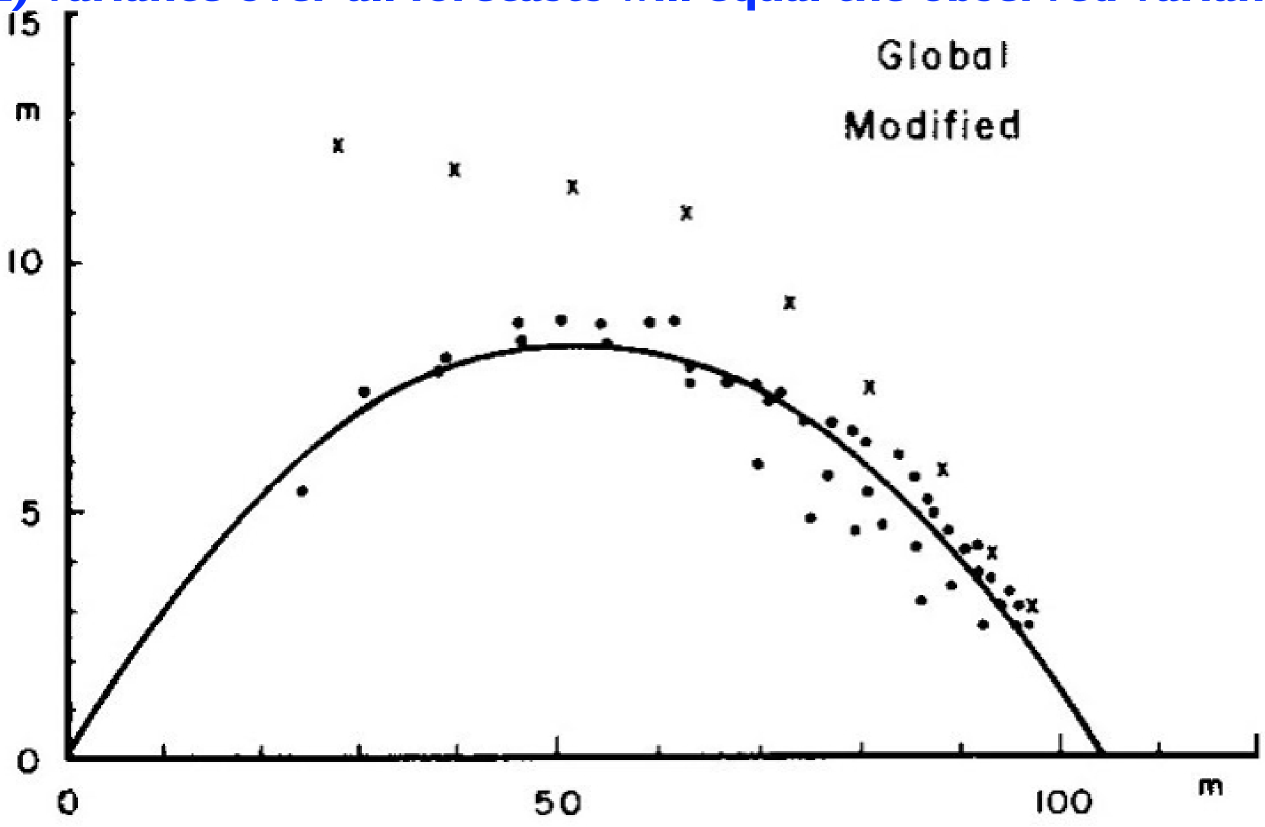


Fig. 3. Same as Fig. 2, but for the modified ECMWF model.

Modified model has smaller error doubling time: 2.16 days

Hemispheric modified model: **this measures the rms Z 500 error for the Northern Hemisphere only, with the statistical correction. Since restricting the analysis to the NH emphasizes the more active winter synoptic systems, we would expect an even lower doubling time (higher error growth) . Here the doubling time is 1.85 days**

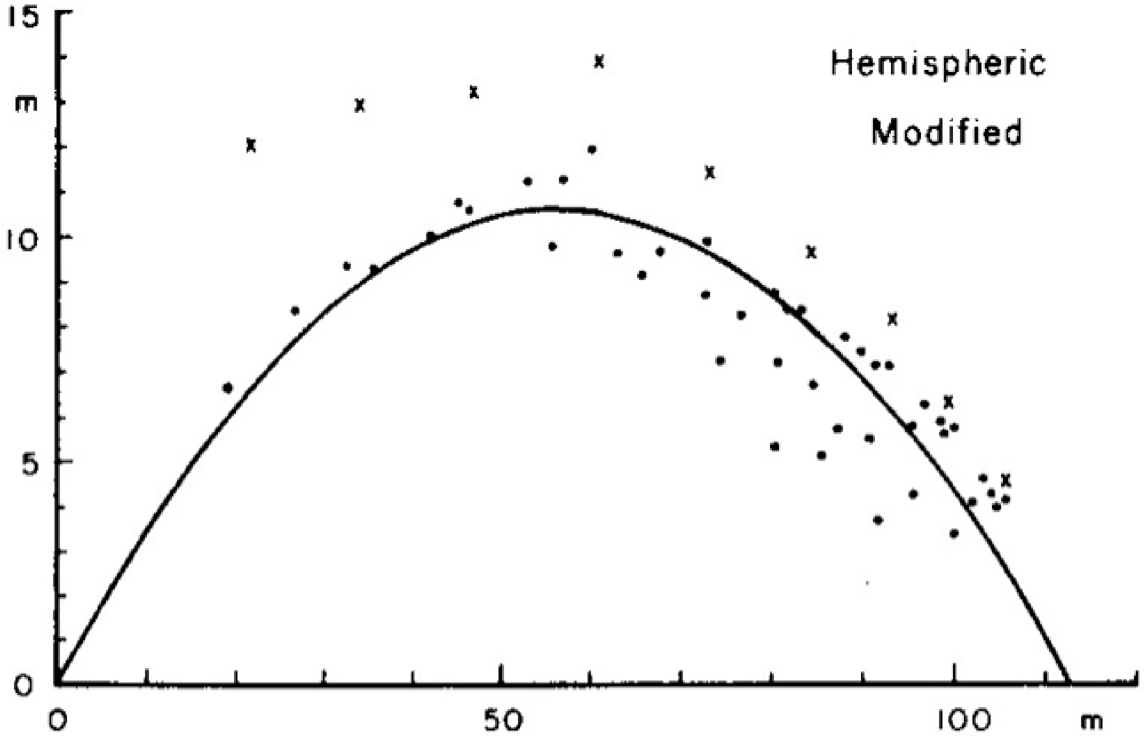


Fig. 4. Same as Fig. 2, but for the modified ECMWF model, for the Northern Hemisphere only.

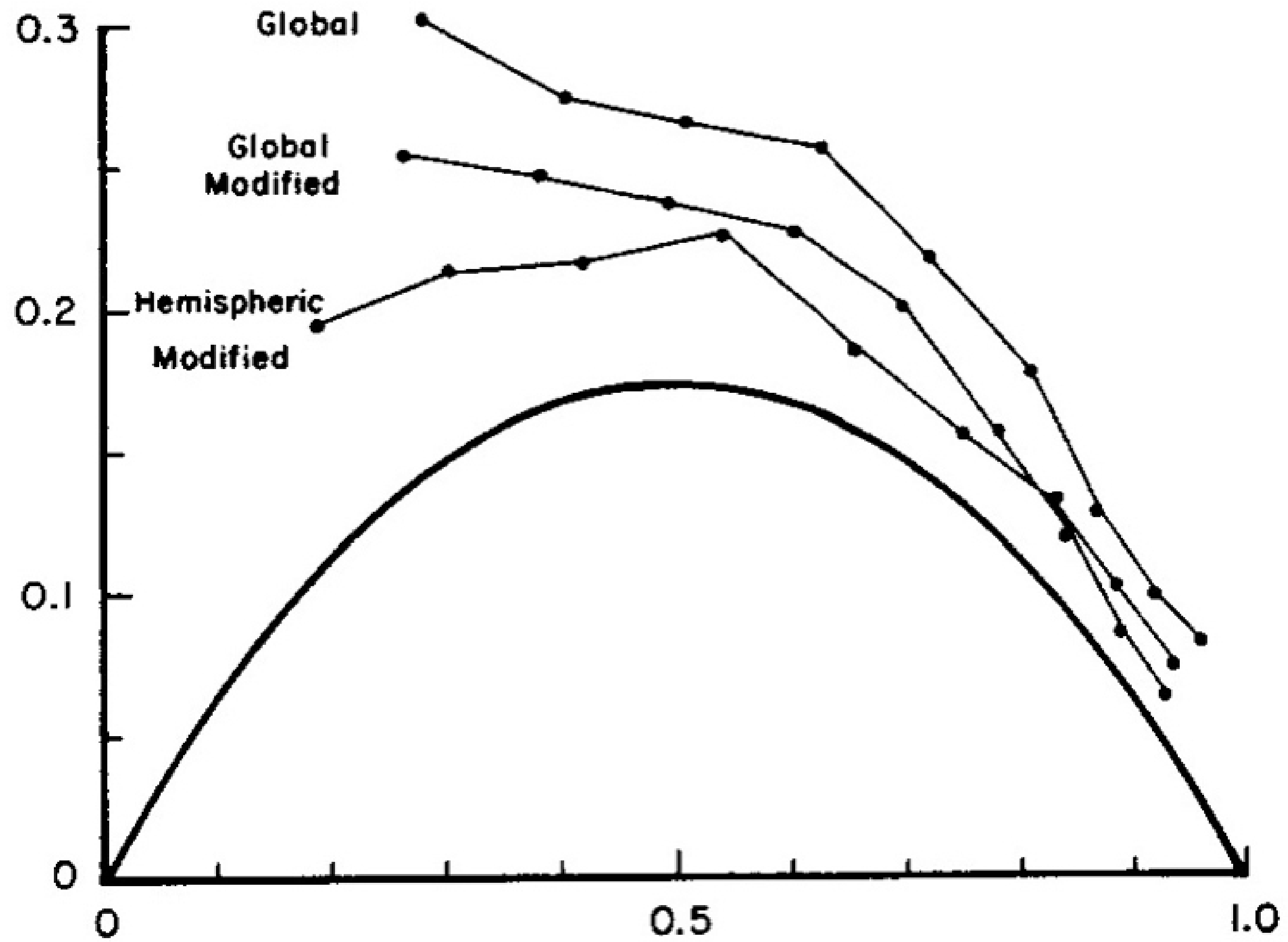


Fig. 5. Superposition of points marked by crosses in Figs. 2–4, after horizontal and vertical scales have been altered so that parabolas coincide. Curves labelled "global", "global modified", and "hemispheric modified" connect points from Figs. 2, 3 and 4 respectively.

Overview of results for winters of 1981/82 - 2000/01

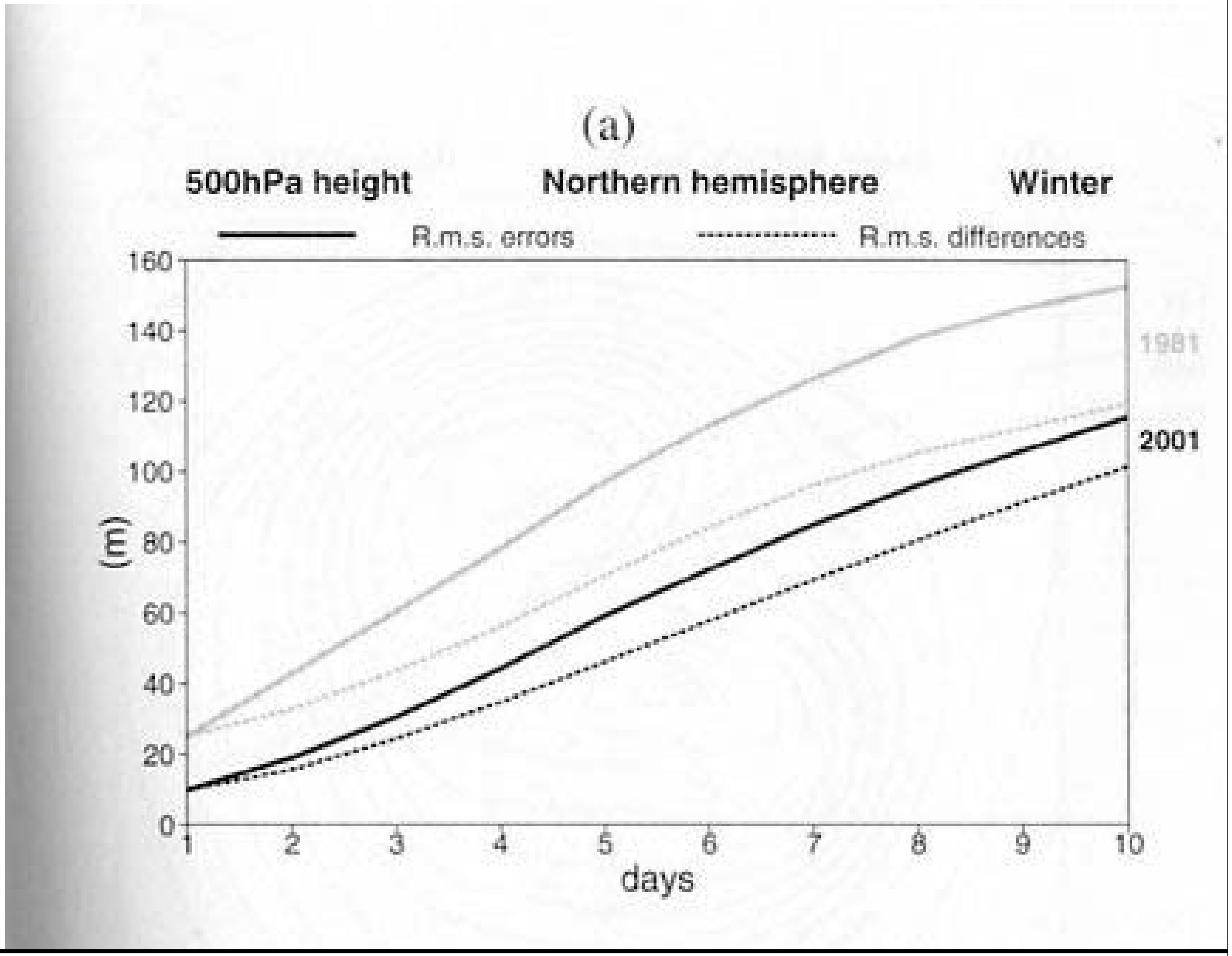
Some general questions that we may ask:

Is the error at day 1 decreasing as we improve the assimilation and models?

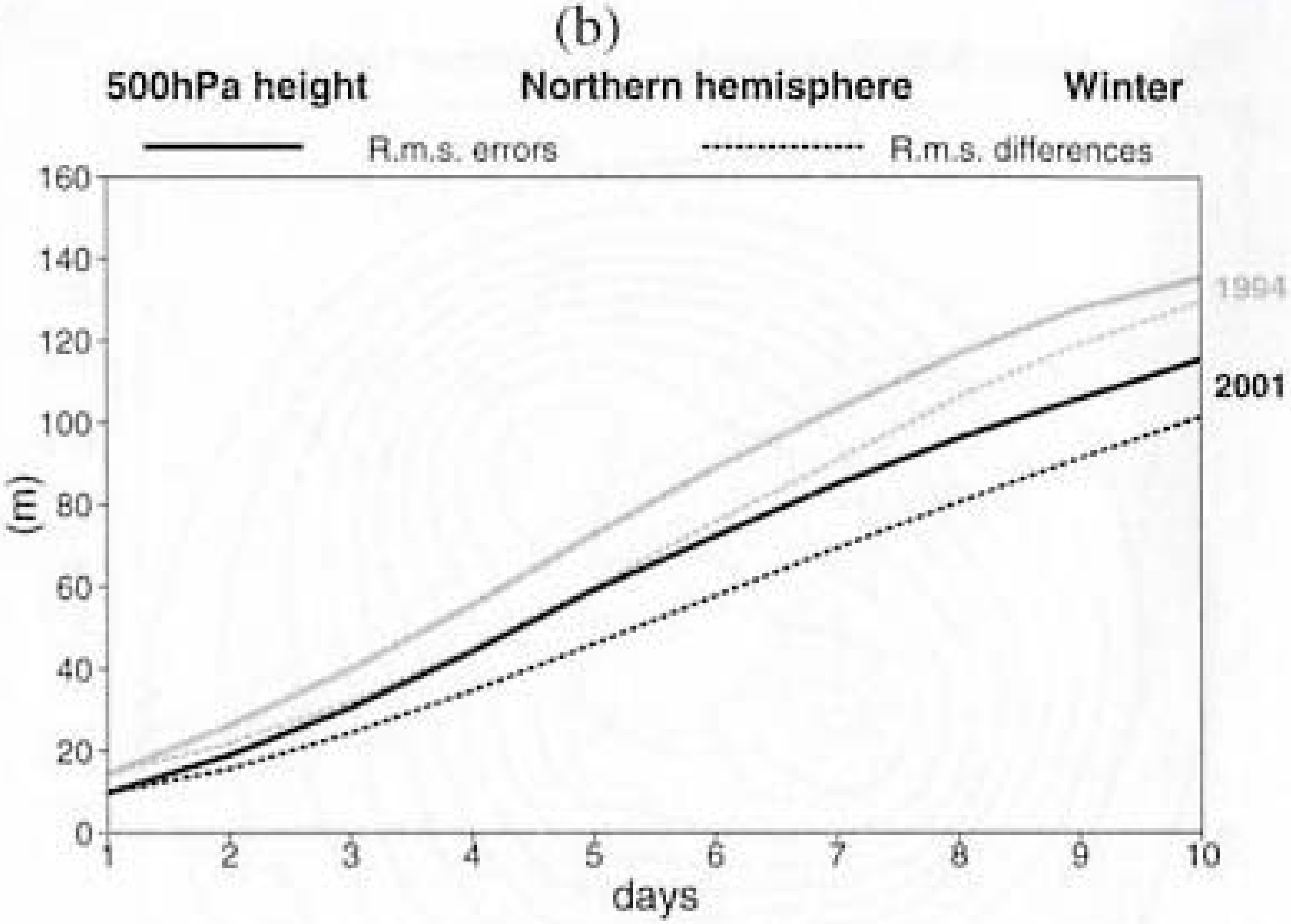
Has the rate of growth of small errors increased or decreased?

Are the “upper limit” and “lower limit” curves getting closer?

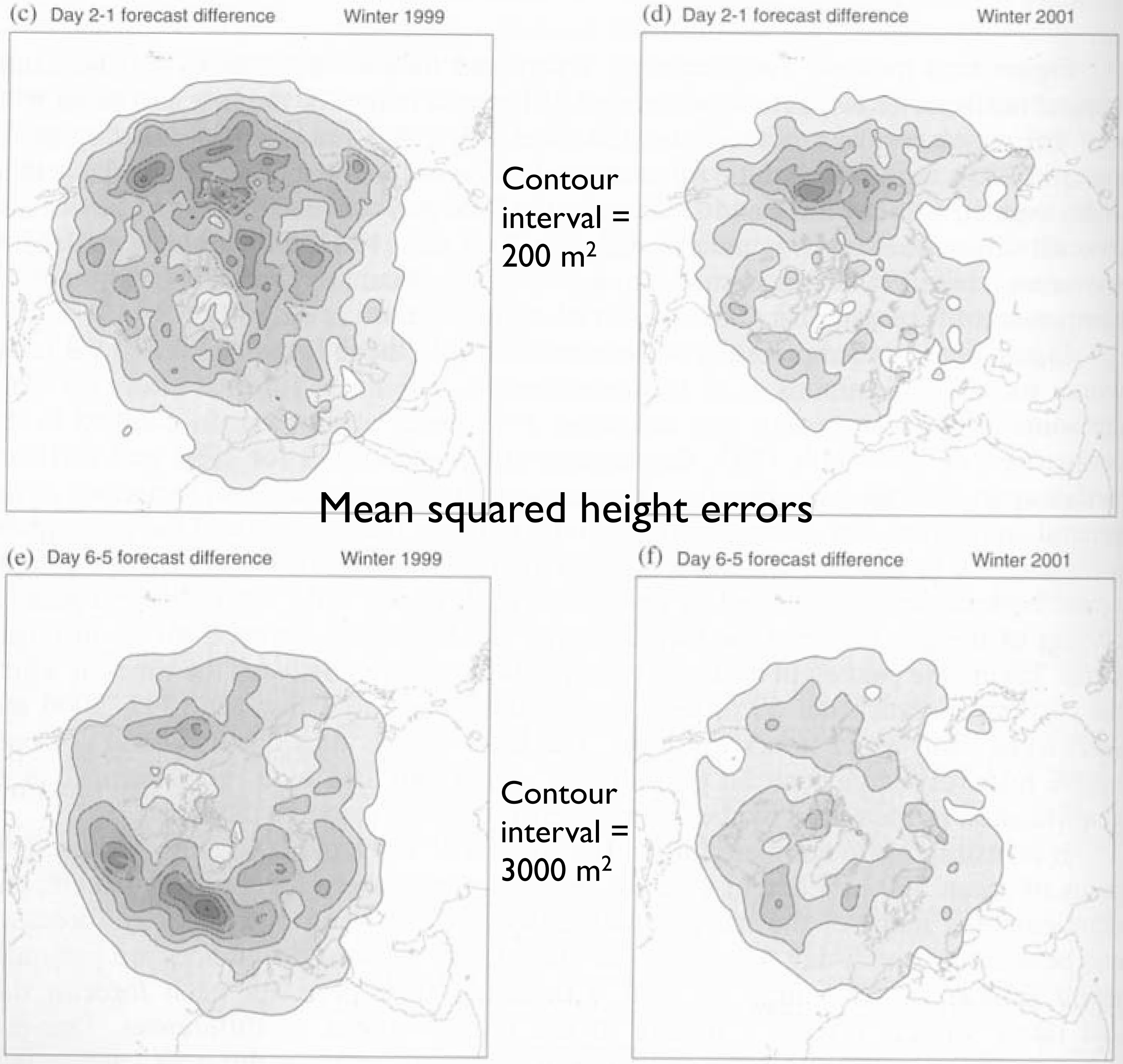
How much does this depend on the dynamics of the particular winter?



Update from 1981 (light curves) to 2001 (dark curves)
- Solid curve is again forecast error.
- Dashed curve is error of forecasts one-day apart valid for same time



Update from 1994 (light curves) to 2001 (dark curves)
- Solid curve is again forecast error.
- Dashed curve is error of forecasts one-day apart valid for same time

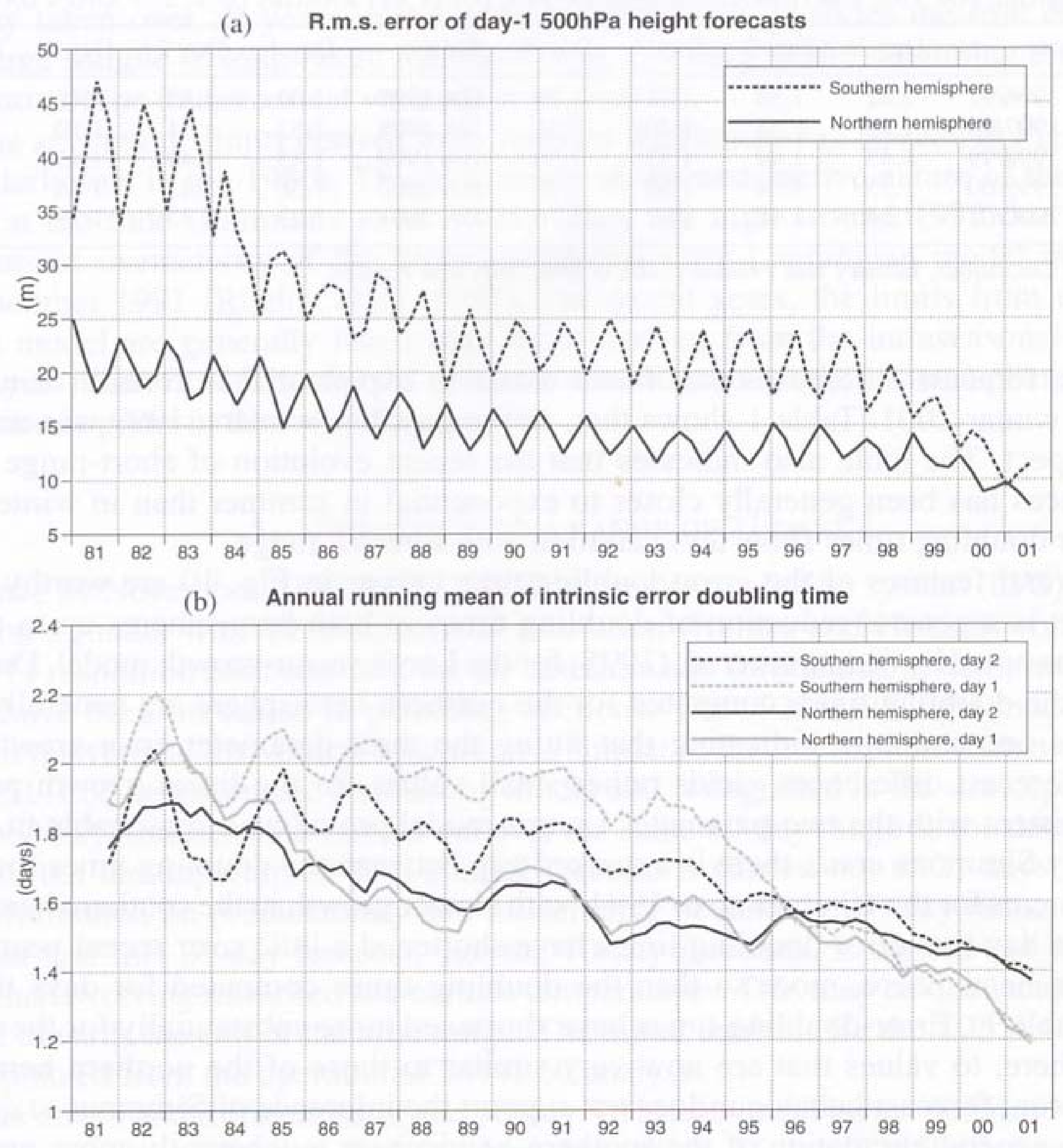


Contour interval = 200 m^2

Mean squared height errors

Contour interval = 3000 m^2

History of errors as the NWP models and analysis systems have improved



An important point to note:

As the error itself has decreased with improved models and better data analysis, the error growth has increased - that is the doubling time has decreased!

We saw this also in the Lorenz analysis - the statistically corrected model had a smaller doubling time.

In this case, the cause of the increased error growth is that the horizontal and vertical resolution of the models has increased over time, so that smaller and smaller scales are resolved. Errors on these smaller scales grow more rapidly!

Solution of three-term error growth model

We rewrite the model in a slightly different form, following Kalnay and Dalcher:

$$\frac{dE}{dt} = (AE + S)\left(1 - \frac{E}{E_\infty}\right)$$

where E_∞ is the value of E as $t \rightarrow \infty$

and S is *hypothesized* by KD to be related to GCM error

with $\frac{e}{E_\infty} = \varepsilon$

$\frac{S}{E_\infty} = \sigma$ **Note that σ is a growth rate (units of inverse time)**

we have:

$$\frac{d\varepsilon}{dt} = (A\varepsilon + \sigma)(1 - \varepsilon)$$

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = \frac{1}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt} = \frac{1}{(1-\varepsilon)}(A\varepsilon + \sigma)$$

$$\frac{df}{dt} = A\frac{\varepsilon}{(1-\varepsilon)} + \sigma\frac{1}{(1-\varepsilon)} = Af + \sigma(1+f) = (A+\sigma)f + \sigma$$

with: $f = \frac{\varepsilon}{(1-\varepsilon)}$ $\quad \varepsilon = \frac{f}{(1+f)}$ $\quad 1 + f = \frac{1}{1-\varepsilon}$

The solution is given by the sum of the general homogeneous solution plus the particular solution:

$$f = Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}$$

$$\varepsilon = \frac{Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}}{1 + Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}}$$

$$\varepsilon = \frac{Ce^{(A+\sigma)t} + \frac{A}{A+\sigma} - 1}{Ce^{(A+\sigma)t} + \frac{A}{A+\sigma}}$$

$$\varepsilon = 1 - \frac{1}{Ce^{(A+\sigma)t} + \frac{A}{\sigma+A}} = 1 - \frac{A+\sigma}{A+(\sigma+A)Ce^{(A+\sigma)t}}$$

where C can be related to the initial error. If the term S = 0, we just obtain:

$$\varepsilon = 1 - \frac{1}{1+Ce^{(A+\sigma)t}}$$

Predictability of Weather
Large Atmospheric Models
Part IIIa

Scale Dependence of Error Growth in Large Models

Zonal Fourier Spectrum:

An Introduction to Scale Selection

Consider any atmospheric field at a pressure level which lies above the surface at all longitudes.

For a fixed latitude, this can be represented simply as a one-dimensional function which is periodic in longitude λ:

$$F(\lambda) = \sum_{m=0}^{m=\infty} A_m \cos(m\lambda) + \sum_{m=1}^{m=\infty} B_m \sin(m\lambda)$$

$$= Re\left[\sum_{m=-\infty}^{m=\infty} C_m e^{im\lambda}\right] \qquad (1)$$

The coefficients A and B are real, while C is complex.
The integer *m* is known as the *zonal wave number*.
m=0 is just the longitudinal average or “zonal mean”.

A Note on the upper limits of the wavenumber m:

In principle there is no upper limit in the expansions because here the function F is continuous.

However, most energy in the real atmosphere is contained in the range m < 20, although the upper limit of m retained in current weather prediction models is as high as ~500

m = 0 - 3 are often called “planetary waves”
m = 0 - 10 are often called “large scale waves”
m = 6 - 20 are often called “synoptic scale waves”
m = 20 -40 are often called “sub-synoptic waves”

From equation (1) we have:

$$A_0 = C_0$$
$$A_m = 2C_m^R \qquad (2)$$
$$B_m = -2C_m^I$$

where R and I denote real and imaginary parts.
To invert the expansion, and obtain the coefficients C, multiply equation (1) on the left by $e^{-il\lambda}$ and integrate over **λ** (where l also refers to zonal wavenumber) :

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} d\lambda\, F(\lambda) e^{-il\lambda} = \sum_{m=-\infty}^{\infty} C_m \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} d\lambda\, e^{im\lambda} e^{-il\lambda} \right)$$
$$= C_l \qquad (3)$$

In the case of a discrete grid of N=2n points around the latitude circle we have:

$$C_l = \frac{1}{N}\sum_{j=-n}^{n-1} F_j e^{-il\frac{2\pi j}{N}}$$

where $\lambda_j = 2\pi j / N$. In terms of A and B:

$$C_0 = A_0 = [F]$$

$$A_m = 2C_m^R = 2\frac{1}{N}\sum_j F_j \cos(m\frac{2\pi j}{N})$$

$$B_m = -2C_m^I = 2\frac{1}{N}\sum_j F_j \sin(m\frac{2\pi j}{N})$$

where we have used the conventional bracket notation to define the zonal mean.

Synthesizing the discrete field F in term of components gives:

$$F_j = [F] + \sum_{m=1}^{M} A_m \cos(m\lambda_j) + \sum_{m=1}^{M} B_m \sin(m\lambda_j)$$

where the largest wavenumber M = n.

We can also write this expansion in terms of the amplitude (**α**) and phase (**ψ**) of different zonal waves:

$$F_j = [F] + \sum_{m=1}^{M} \alpha_m (\cos(m\lambda_j - \Psi_m))$$

$$\alpha_m = \left(A_m^2 + B_m^2\right)^{\frac{1}{2}}$$

$$\tan(\Psi_m) = B_m / A_m$$

Note that the tangent is to be computed from the point (A_m,B_m) in the polar plane.

It is straightforward to prove that, neglecting the zonal mean, we have:

$$[F^2] = \frac{1}{2} \sum_{m=1}^{M} \left(A_m^2 + B_m^2\right)$$

If F_1 and F_2 are two fields, the squared "error" $[(F_1 - F_2)^2]$ = $[(\delta F)^2]$ can be written as:

$$[(\delta F)^2] = \frac{1}{2} \sum_{m=1}^{m=M} \left((\delta A_m)^2 + (\delta B_m)^2\right)$$

An interesting way to write the squared error *for a single wavenumber* is in terms of the amplitude (α) and phase (Ψ) of each solution (denoted by subscripts 1 and 2):

$$\left[\delta F^2\right] = \frac{1}{2}\sum_{m=1}^{M}\left[(\alpha_1 - \alpha_2)^2 + 2\alpha_1\alpha_2\left(1 - \cos(\Psi_1 - \Psi_2)\right)\right]$$

where all terms inside the square brackets depend on m. Clearly the first term is the error due to the amplitudes of the waves differing (it vanishes when the amplitudes are the same), and the second term is the error due to differing phases (it vanishes when the phases are the same).

Weather Predictability - Part IIIb
Scale Dependence of Error Growth

David M. Straus

George Mason University

Center for Ocean-Land-Atmosphere Studies (COLA)

Dan Paolino

Center for Ocean-Land-Atmosphere Studies (COLA)

Overview of Global Predictability

- *Mid-latitude* weather error growth dominated by baroclinic instability:
  - Synoptic scales to large scales involved
  - Time scales of 1-10 days (very roughly)
  - Release of latent heat (condensation of water) plays a secondary role

- *Tropical* weather error growth involves processes on several scales:
  - Deep convection (latent heat release plays dominant role)
    - Very small spatial scales (~1 km)
    - Short time scales (hours)
  - Tropical Waves and the Madden-Julian Oscillation
    - Synoptic to planetary scales
    - Range of time scales 1 - 60 days

Identical Twin Experiment

- **Simulations using an atmospheric model (AGCM)**
  - **Horizontal Resolution T63 (maximum of 63 zonal waves)**
  - **forced by observed SSTs (given each week) for winter only**

- **"Control Experiments": For each of 18 winters (1981-82 / 1998/99)**
  - **10 "control" simulations started from analyses in Late November: (Nov. 30, 29, 28, … ,21)**
  - **Each simulation run for at least 60 days**

- **"Perturbed Experiments": For each of 18 winters (1981-82 / 1998/99)**
  - **10 "perturbed" simulations started from analyses in Late November: (Nov. 30, 29, 28, … ,21) but with added, very small, random, errors*.**
  - **Each simulation run for at least 60 days**

Identical Twin Experiment (continued)

- *Error in each variable (for example T) is given by dT = T * 0.001 * r, where r is a random number in the range (-1,1)

- Such errors (or perturbations) are applied globally.

- Each Control Experiment thus has a "twin" Perturbed Experiment:
  - They start from almost the same initial condition
  - They are forced by the same, time varying, SSTs.

- The error is defined as the difference between a Control Experiment and its twin.

- The error variance is defined as the error^2, averaged over longitude, and then averaged over all 18 x 10 = 180 identical twin pairs.

Variance and Error Variance - The Ergodic Hypothesis

There are two ways defining a climatological mean in general from a set of atmospheric GCM experiments:

(1) Average over all possible model states consistent with winter. This is called an *ensemble* average, and can be obtained by sampling states from a number of simulations.

(2) Time average over a single winter simulation that is long enough - for example a very long simulation of the model using constant winter SSTs.

The equivalence of these two types of averages, ensemble average and time average, is called the ergodic hypothesis. The point is that if we follow a single solution (trajectory) of the model for a long enough time, it will eventually cover all the states of the system consistent with winter boundary conditions.

Variance and Error Variance: Saturation at for Large Forecast Time

Let $<T>$ be the ensemble average defined in the previous slide. Then the *variance* V is defined as:

$$V = < (T - <T>)^2>$$

In practice this is estimated by taking the temporal variance of T about its seasonal mean for a given winter simulation, and averaging over all simulations. It does not depend on time.

The *error variance* E is defined as:

$$E = <(T_{(1)} - T_{(2)})^2>$$

where (1) and (2) refer to pairs of twins. This clearly has a sense of time, since it is very small at small forecast time, and becomes very large as time increases.

But we have:

$$E = <(T_{(1)} - T_{(2)})^2> = < ((T_{(1)} - <T>) - (T_{(2)}-<T>))^2> =$$

$$< (T_{(1)} - <T>)^2> + < (T_{(2)}-<T>) ^2> - 2 < ((T_{(1)} - <T>)(T_{(2)}-<T>))>$$

The last term is just the covariance between the two twin forecasts. For large enough forecast time, we expect the covariance to approach zero, since the forecasts are expected to be completely uncorrelated.

Variance and Error Variance: Saturation at for Large Forecast Time

In general, then we have

$$E = < (T_{(1)} - <T>)^2> + < (T_{(2)}-<T>)^2> - 2 < ((T_{(1)} - <T>)(T_{(2)}-<T>))>$$

At large times, the last term is expected to vanish. Each of the first two terms is a valid estimate of the variance, so that we have the result:

For large times, the error variance becomes twice the variance:

$$E = (t \rightarrow \infty)\ 2\ V$$

This value is called the *saturation* value - Once the error variance has reached this value it is saturated, and will not grow any further.

The error variance divided by the saturation value is called the normalized error variance E_N:

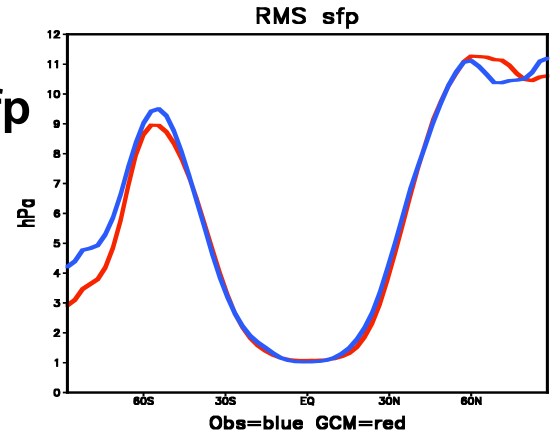
$$E_N = E/(2V) = (t \rightarrow \infty)\ 1$$

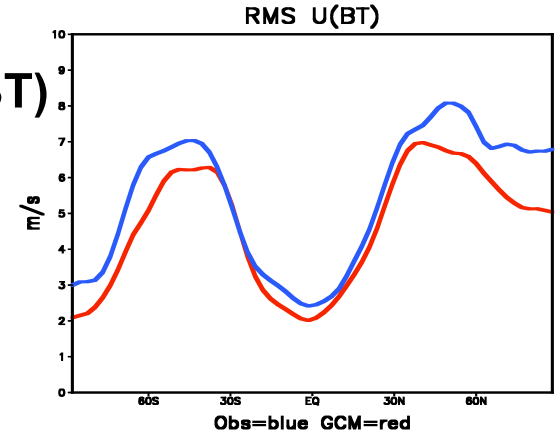
The normalized error variance saturates at the value of 1

Please note that this applies only to the ensemble averaged error variance, so that the brackets < > refer to an average over many pairs of twins.

Square root of variance

Obs **GCM**

sfp


u(BT)


sfp = surface pressure

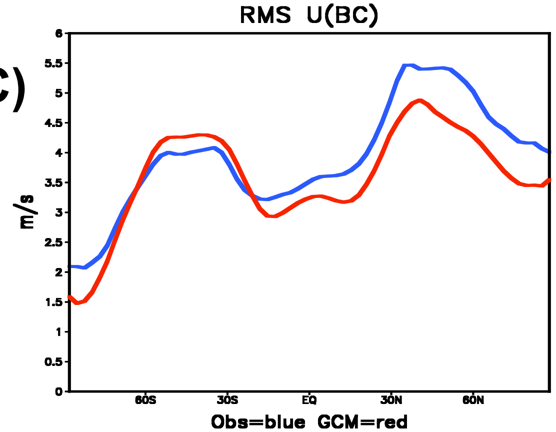
u = zonal wind

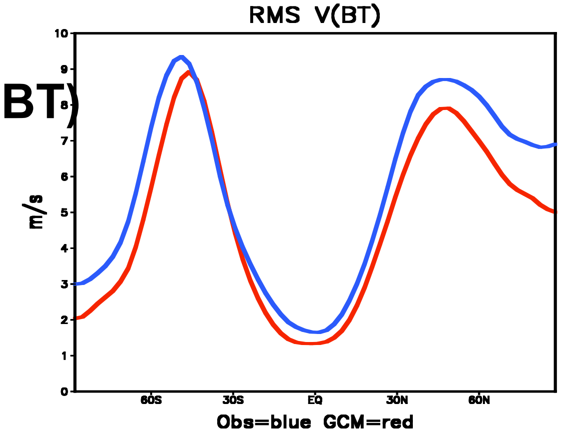
v = meridional wind

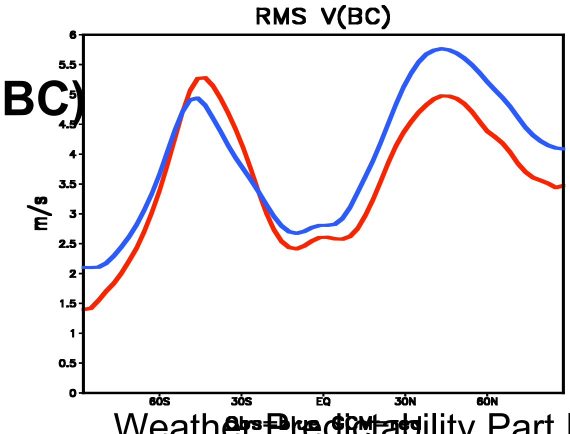
q = specific humidity

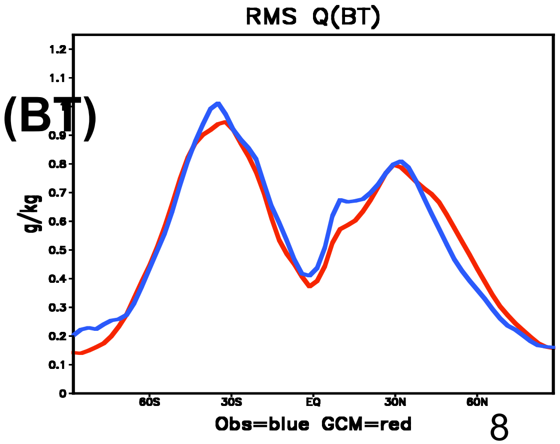
BT means vertical average

BC means deviation from vertical average

u(BC)


v(BT)


v(BC)


q(BT)


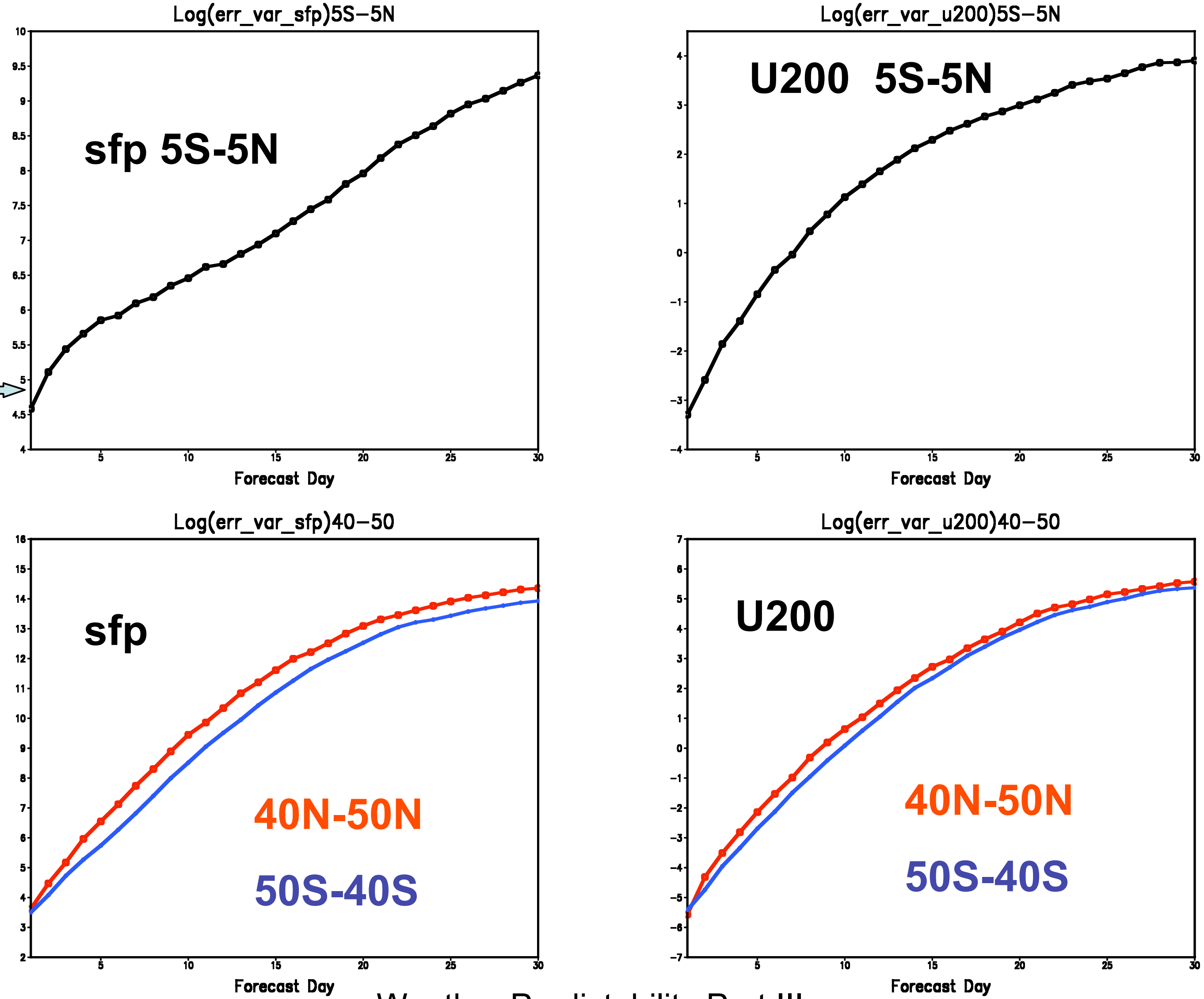
Tropical variance much less than mid-latitude variance:

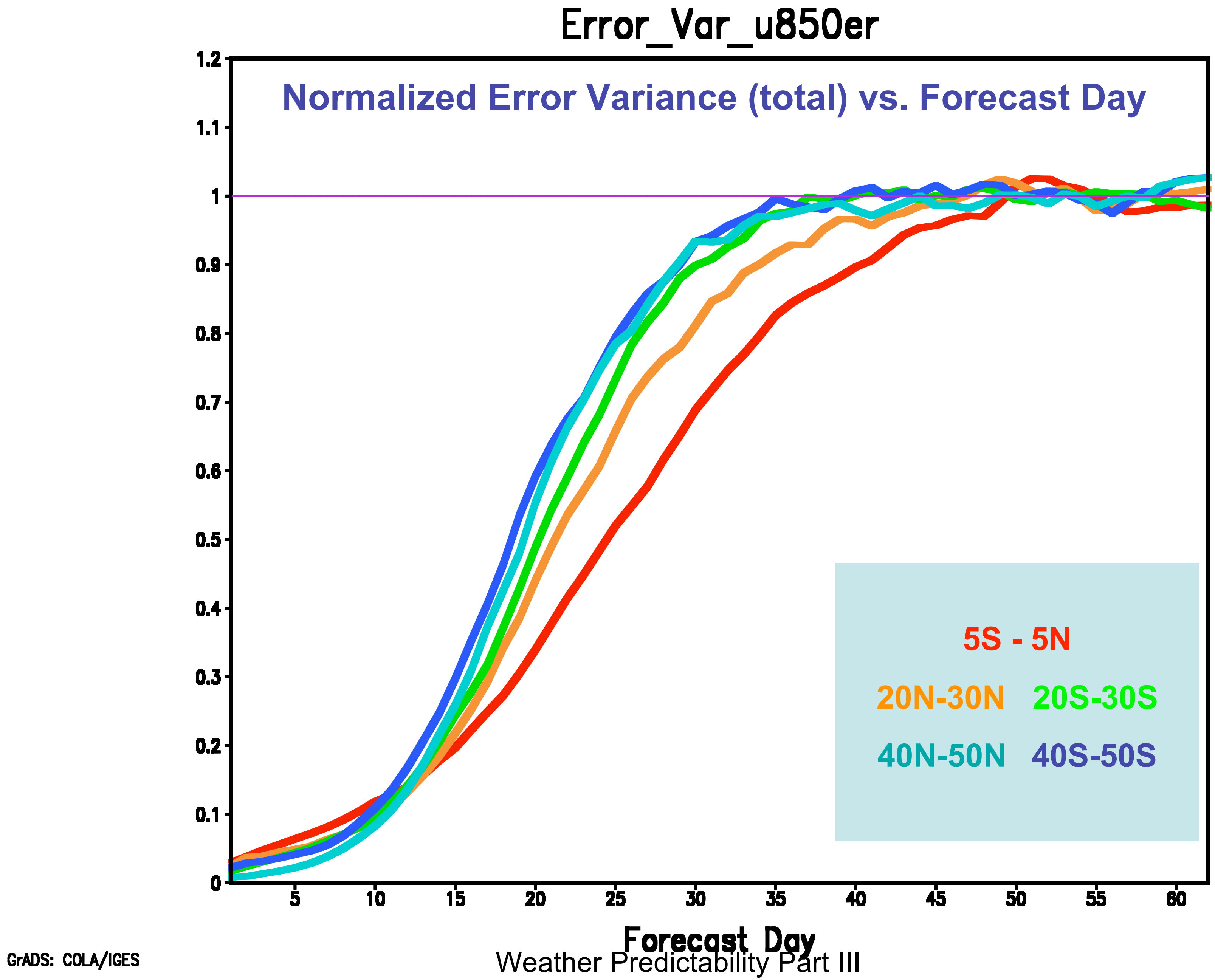
Expect smaller values of E

$$\log(E) = \alpha + \beta t \quad \Rightarrow \quad E = E_0 e^{(\beta t)}$$

Log (error) vs. Forecast Day

Note rapid error growth for very small time!




Error_Var_u850er
Normalized Error Variance (total) vs. Forecast Day
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50
55
60
5S - 5N
20N-30N
20S-30S
40N-50N
40S-50S
Forecast Day
GrADS: COLA/IGES

(1) For short times, mid-latitude error growth slower than tropical error growth:

convective instability sets in very rapidly compared to baroclinic instability.

(2) For longer times, mid-latitude error growth is more rapid as baroclinic instability sets in.

(3) Mid-latitude error saturates at higher values than tropical error, but you can't see that using normalize error!

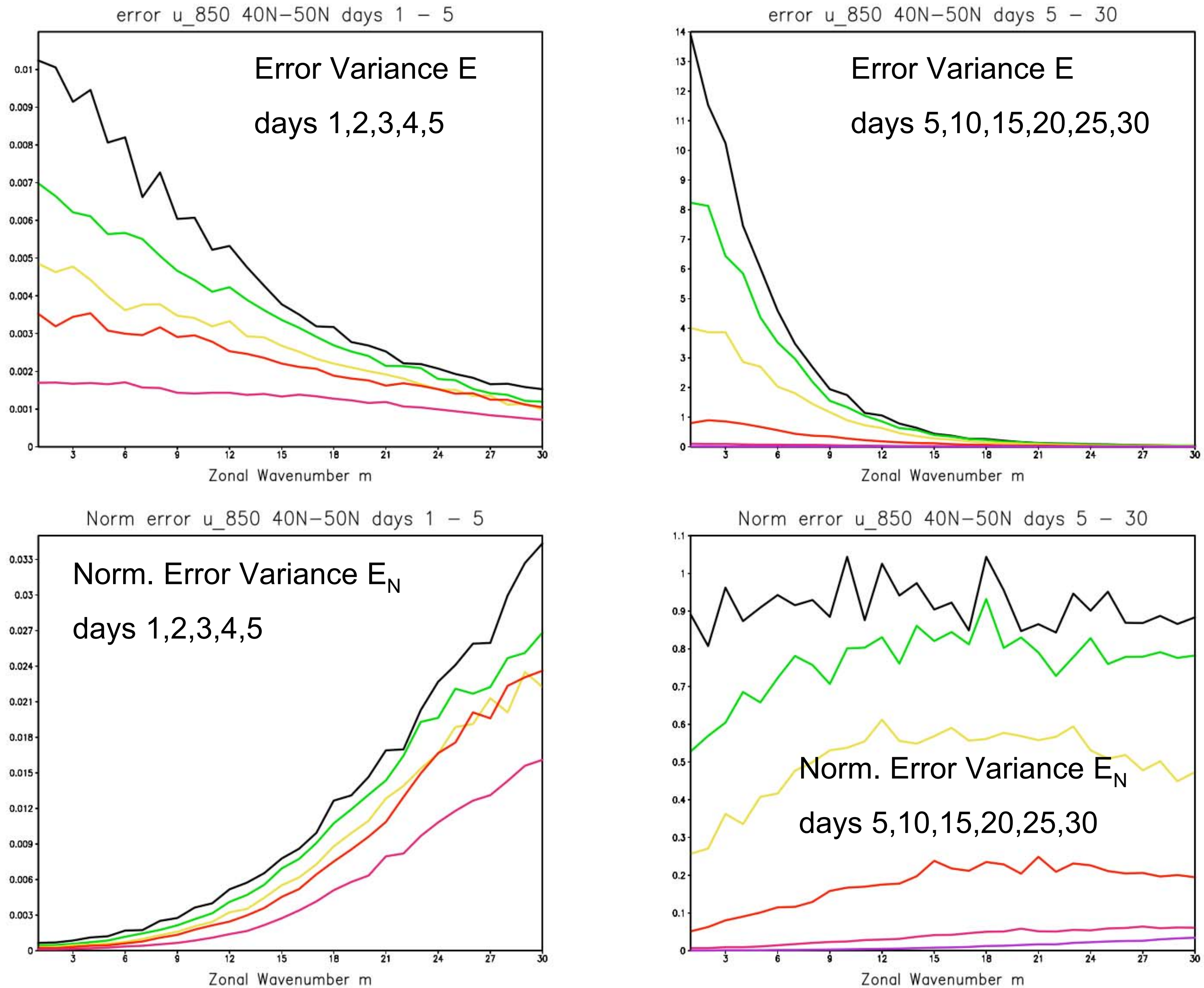
(4) Exact time for errors to saturate is hard to estimate -so:

***the predictability time* τ is estimated as the time it takes for E_N to reach 1/2.**

u 850 hPa

E and E_N

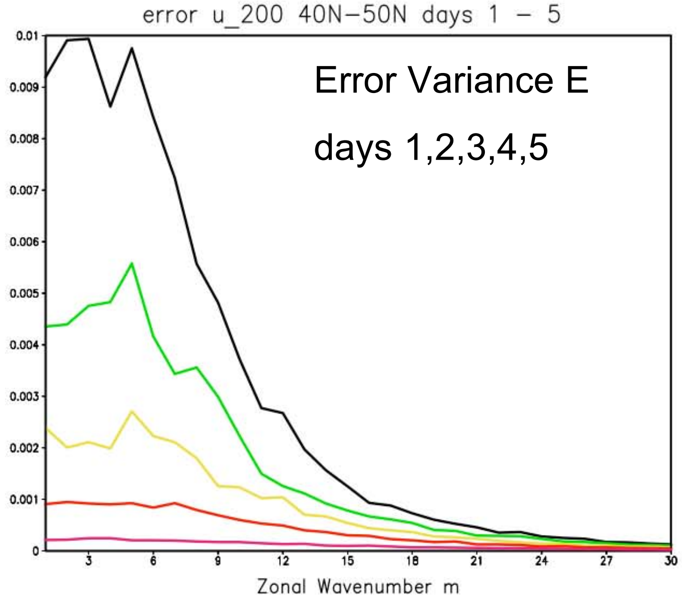
40-50N

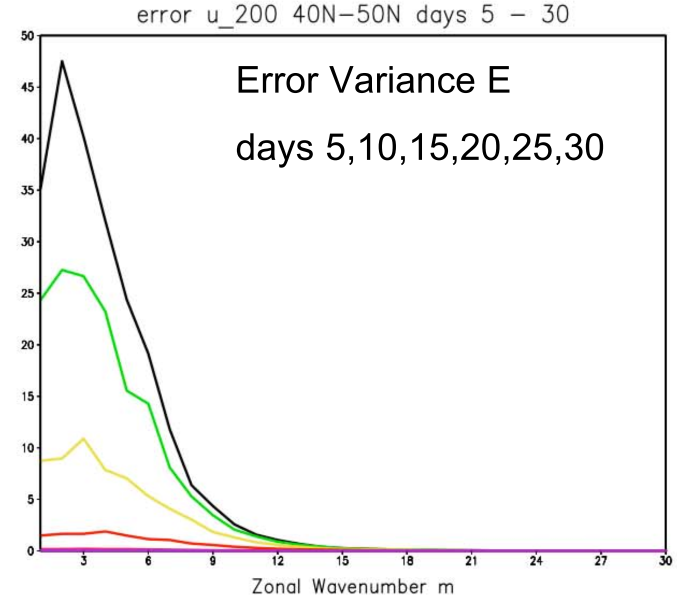


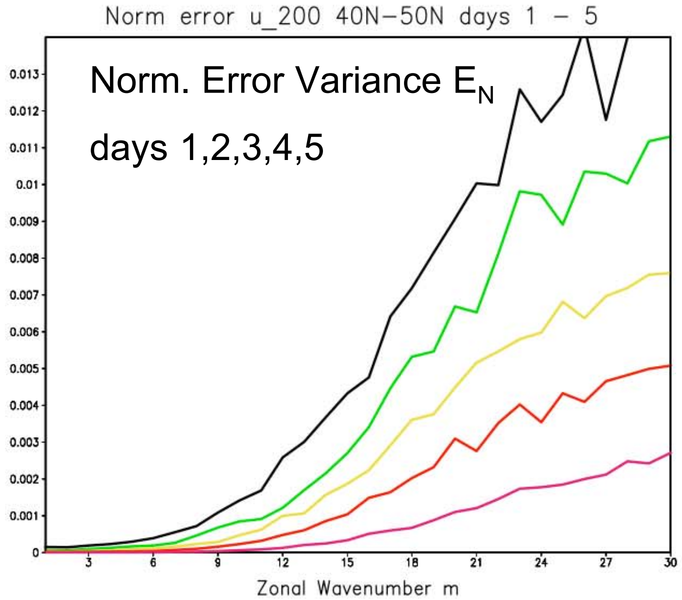
u 200 hPa

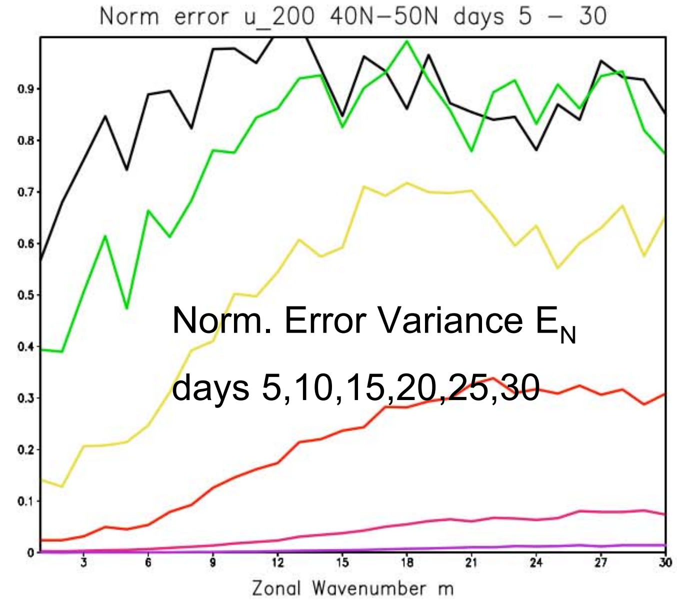
E and E_N

40-50N









(1) The error E initially grows more rapidly at large scales (smaller values of m) than at smaller scales (larger values of m)

(2) But this is because there is more variance at large scales (in mid-latitudes)

(3) Once E is divided by the variance to obtain E_N, we see that the smaller scales grow more rapidly - that is they approach their saturation values more quickly than do the larger scales

(4) Another way of saying this is that the predictability time τ is longer for large scales-This is very apparent at 200 hPa, less so at 850 hPa

(5) *Even though the error at large scales is larger than at small scales, the large scales in the pairs of identical twins do not become completely de-correlated as rapidly as the small scales.* Thus for example, at day 20 at 200 hPa, there is still some useful information left in the large scales (but not in the small scales).

Comparison of tropical vs. mid-latitude error variance growth for different scales

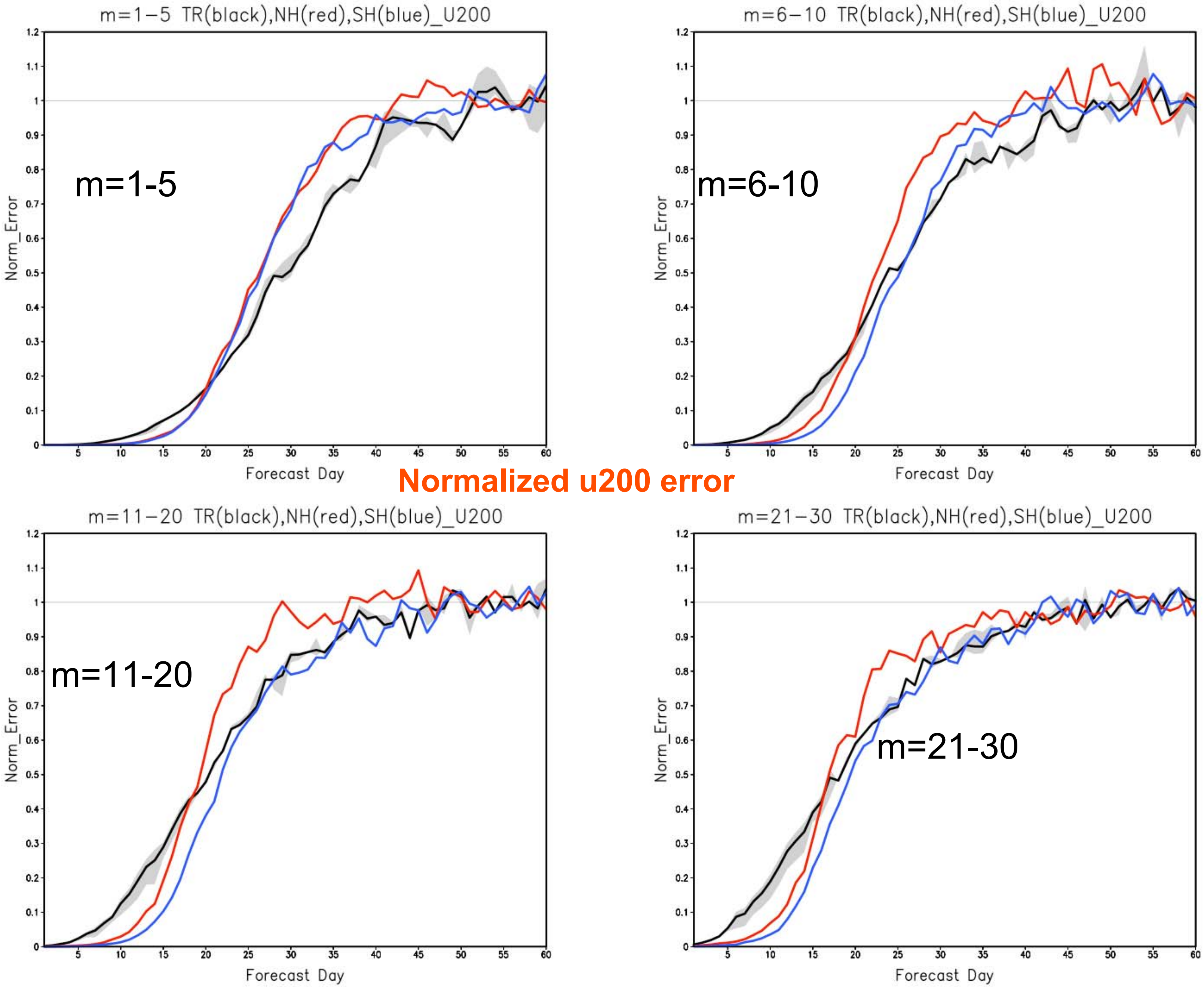
The following plots show normalized error variance growth, summed over wavenumbers m=1-5, m=6-10, m=11-20, and m=21-30, as a function of time

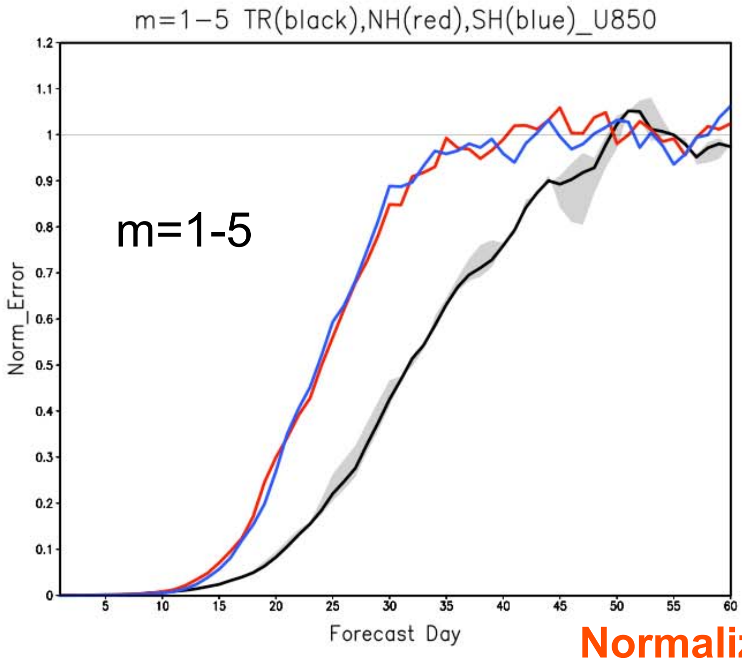
Red lines show results averaged over **Northern mid-latitudes 40N-50N**

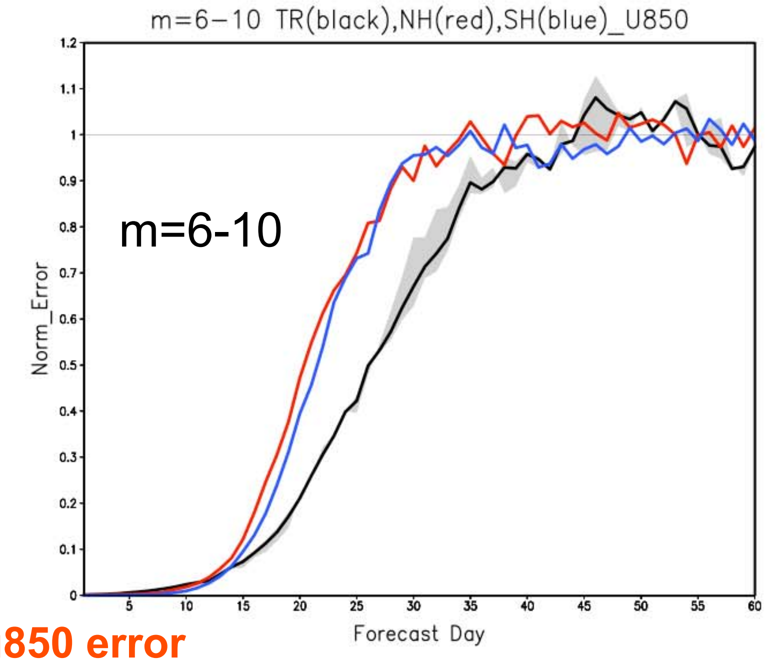
Blue lines show results averaged over **Southern mid-latitudes 40S-50S**

Black lines show results averaged over the **deep tropics (5S-5N).**

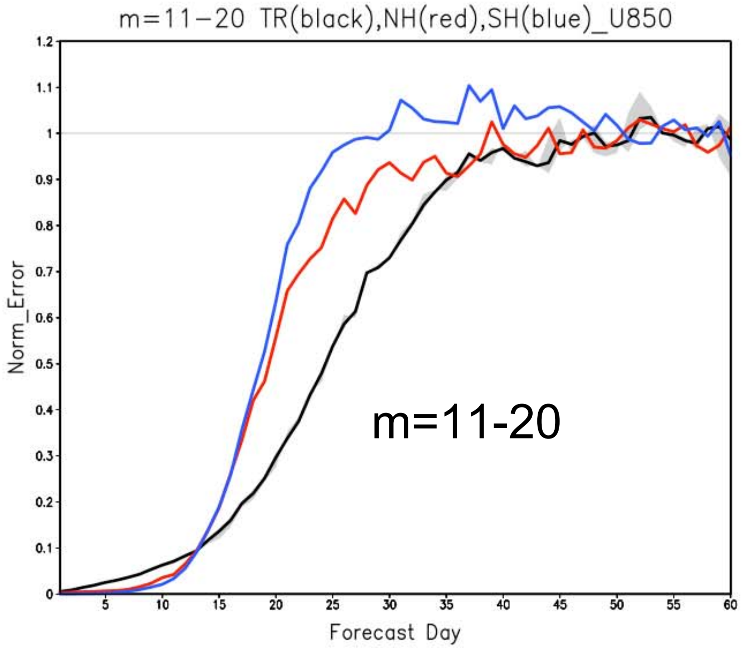
(The grey shaded band around the tropical error curve gives the estimate of the variability from year to year of the error growth).

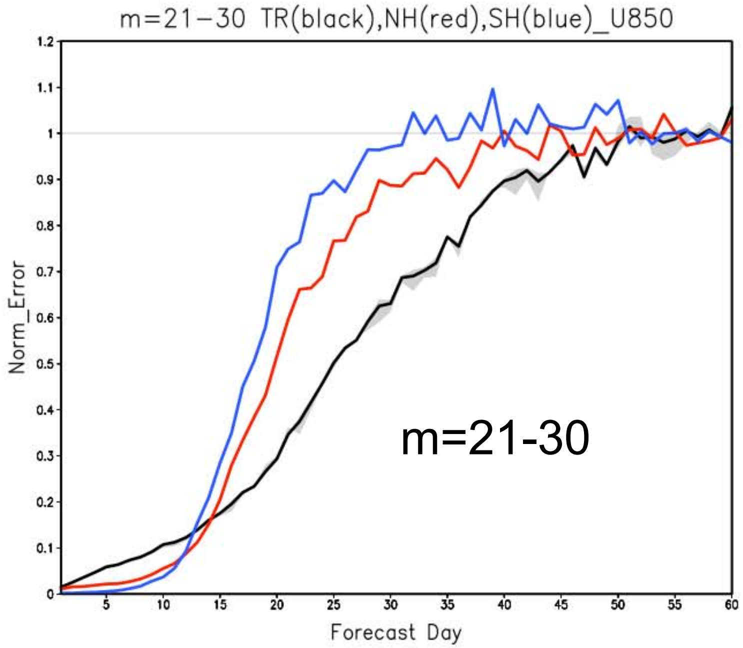

m=1-5 TR(black),NH(red),SH(blue)_U200
m=1-5
m=6-10 TR(black),NH(red),SH(blue)_U200
m=6-10
Normalized u200 error
m=11-20 TR(black),NH(red),SH(blue)_U200
m=11-20
m=21-30 TR(black),NH(red),SH(blue)_U200
m=21-30
Norm_Error
Forecast Day





Normalized u850 error





Comparison of tropical vs. mid-latitude error variance growth for different scales

The following plots show normalized error variance growth, summed over wavenumbers m=1-3, m=4-6, m=7-9, and m=10-12, as a function of time

Red lines show results averaged over **Northern mid-latitudes 40N-50N**

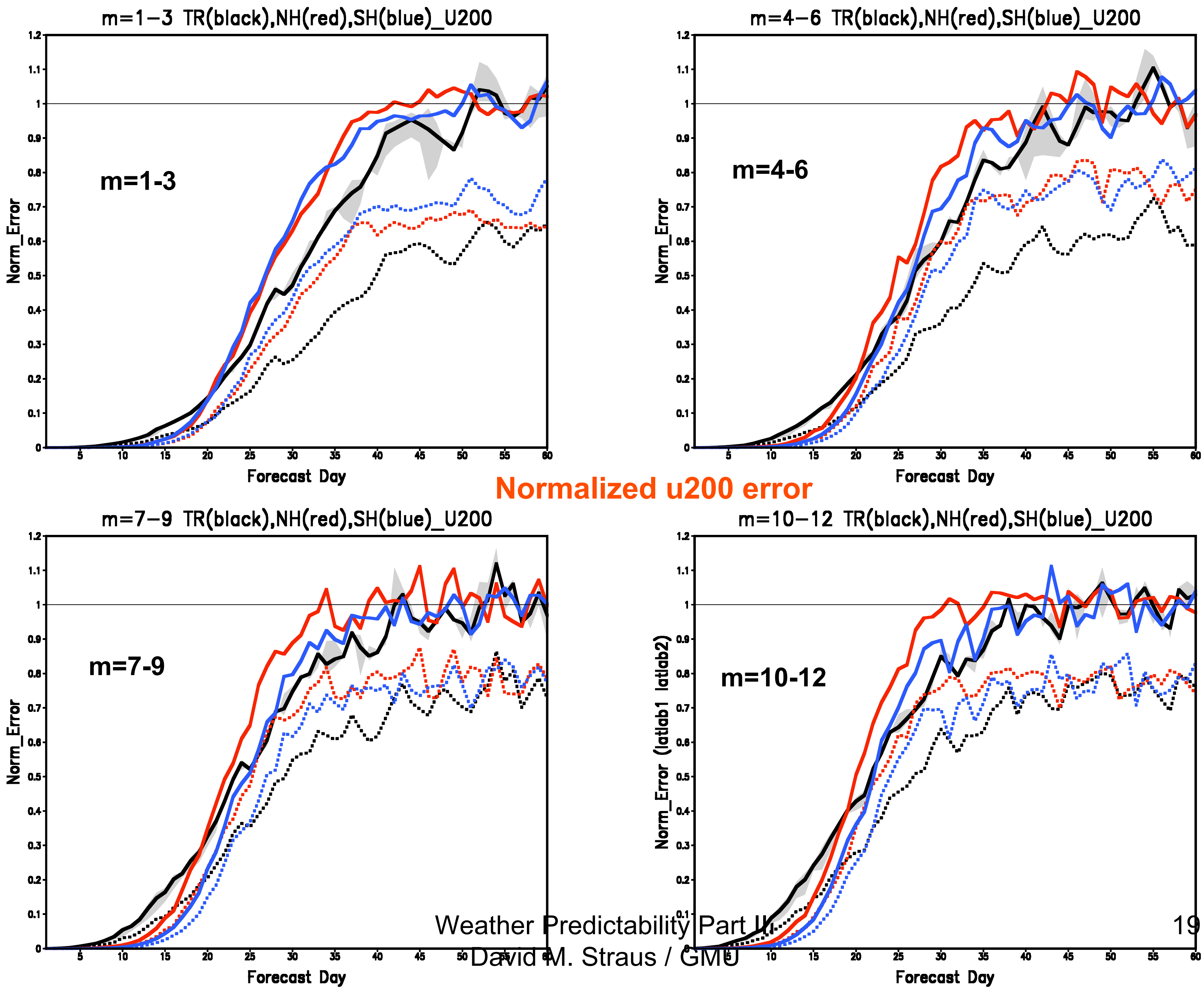
Blue lines show results averaged over **Southern mid-latitudes 40S-50S**

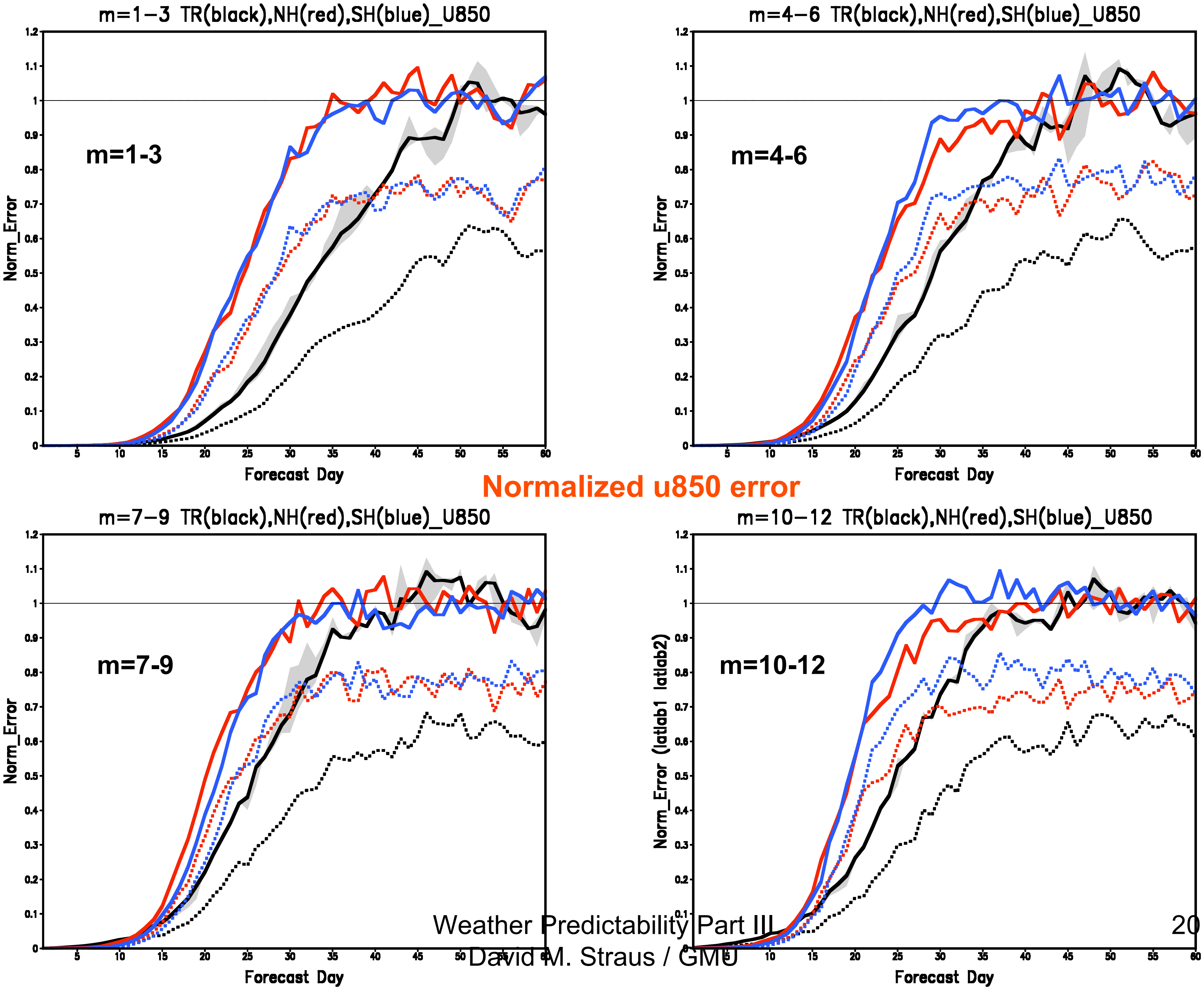
Black lines show results averaged over the **deep tropics (5S-5N).**

(The grey shaded band around the tropical error curve gives the estimate of the variability from year to year of the error growth).

The dotted lines show the growth of errors due to phase differences only (see Lecture IIIa)

Normalized u200 error




m=1−3 TR(black),NH(red),SH(blue)_U850
m=1-3
m=4−6 TR(black),NH(red),SH(blue)_U850
m=4-6
Normalized u850 error
m=7−9 TR(black),NH(red),SH(blue)_U850
m=7-9
m=10−12 TR(black),NH(red),SH(blue)_U850
m=10-12
Norm_Error
Norm_Error (latlab1 latlab2)
Forecast Day

(1) Generally the mid-latitude predictability time is longer for the large scales:

τ ~ 25 days for m=1-5, but τ ~16-18 days for m=21-30.

(2) Generally the tropical error growth is generally slower after day 10 than the mid-latitude error growth; this is seen especially in the largest scales and at lower levels (850 hPa)

(3) The slow tropical error growth at 850 hPa at the largest scales suggests that SST may have a greater effect in regulating the large scales in the tropics than in mid-latitudes.

(4) The phase errors contribute more than half of the total errors.